In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

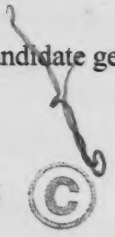
While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

University of Alberta

Characterization of the cat eye syndrome candidate gene *CECR2*

by

Graham Scott Banting



A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta
Fall 2003

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitons et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87933-X
Our file Notre référence
ISBN: 0-612-87933-X

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Graham Scott Banting

Title of Thesis: Characterization of the cat eye syndrome candidate gene *CECR2*

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

9/30/03

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Characterization of the cat eye syndrome candidate gene *CECR2*" submitted by Graham Scott Banting in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular Biology and Genetics.

Dr. H. McDermid

Dr. D. Pilgrim

Dr. A. Underhill

Dr. D. Rancourt

9/26/03

Dr. N. Adames

ACKNOWLEDGEMENTS

I would like to thank Heather McDermid for being an excellent supervisor and giving me a great deal of support throughout the duration of the project. I must also acknowledge the excellent technical contributions of technologist Melanie Kardel, who took over care of the mouse colony once it got too large for me. Without Melanie, this project would surely not have been possible. In must also acknowledge Tanya Ames who took over this project and provided mouse handling/genotyping assistance during the writing of this thesis. Thanks also go out to current and previous members of the McDermid lab for excellent technical and scientific discussions. The sharing of information was always enjoyable.

I must also thank members of the Biological Sciences Department who also contributed greatly to this project. In no particular order, Gary Ritzel for being an excellent source of information and providing stimulating conversation, Jack Scott and Randy Mandryk from the microscopy unit for extremely useful help with microscopy and photography, Dan Bushey for providing PCR primers, and longtime friend Paul Stothard for keeping things in much needed perspective. My committee members for the length of the project David Pilgrim and Rachael Wevrick were also great sources of advice with regards to research directions.

Lastly I would like to thank my friends and family. My parents, and friends back home in Victoria, were always supportive of my strange scientific endeavours, while Lindsay Bridgland was a major source of support throughout many of the ups and downs during my time here.

ABSTRACT

Cat eye syndrome (CES) is caused by a duplication of human chromosome 22q11. CES is characterized by ocular coloboma, anal atresia with or without fistula, down-slanting palpebral features, preauricular tags and/or pits, hypoplastic kidney and congenital heart defects, in addition to other less common features. The goal of this study was to identify and characterize genes within the critical region for CES. Two novel genes were identified, and one (*CECR2*) was chosen for further study. *CECR2* was determined to be a widely expressed gene with a wide variety of alternative isoforms. Sequence analysis suggested that *CECR2* was likely involved in chromatin remodeling and transcriptional regulation of other genes. An ES cell line containing a genetrap within *Cecr2* was used to generate a 'knockout' mouse. Expression of *Cecr2* via the genetrap reporter gene was determined to be predominantly neural. Mice homozygous for the *Cecr2* genetrap developed exencephaly (anencephaly) with moderately high penetrance, suggesting that *Cecr2* is important for neurulation. Transgenic mice were generated which carried a BAC transgene (containing *CECR2*) in order to determine if (over) expression of *CECR2* in mice would produce CES like phenotypes. Two transgenic lines were generated, one appearing normal and one showing embryonic lethality, however it was unclear if the embryonic lethality was related to the expression of genes contained within the transgene. *CECR2* transgenic mice were crossed to *Cecr2* knockout mice to determine if the transgene would rescue the exencephaly phenotype. No rescue was observed suggesting that either the transgene was not producing active protein, or alternatively that the human protein was not functional in mouse. Further characterization will be required to exclude *CECR2* as a CES candidate gene.

TABLE OF CONTENTS

1. INTRODUCTION 1
1.1 Gene dosage 1
1.2 Chromosome 22 1
1.3 Cat eye syndrome (CES) 4
1.3.1 CES history 4
1.3.2 Disorders that share phenotypic overlap with CES 5
1.3.3 Monogenic vs polygenic 7
1.3.4 Delineating the CES critical region (CESCR) 8
1.3.5 Sequence analysis of the CESCR 9
1.3.6 Candidate gene approach 11
1.3.7 Mouse modeling of CES 14
1.4 Transcriptional Co-activators 17
1.4.1 Histone modification and gene expression 19
1.4.2 Co-activator signature motifs 21
1.4.3 Mutations in selected co-activators 23
1.5 Neural tube defects 24
1.5.1 Neural tube closure 24
1.5.2 Mouse models of NTDs 25
1.5.3 Genetics, environment and dietary supplements 27
1.6 *Drosophila* genetic models 30
1.7 Main objectives of this work 31
2. MATERIALS AND METHODS 34
2.1 DNA Isolation 34
2.1.1 Plasmid 34
2.1.2 Mouse genomic 34
2.1.3 BAC DNA purification for pronuclear injection 35
2.1.4 Agarose embedded DNA 35
2.2 PCR and cloning 36
2.2.1 PCR buffer 36
2.2.2 Reverse Transcriptase PCR 36
2.2.3 5'/3' RACE 36
2.2.4 Genomic 'RACE' 37
2.2.5 Single fly PCR 37
2.2.6 CECR2/Cecr2 cloning 38
2.2.7 Expression constructs 38
2.3 Mouse genetic engineering 39
2.3.1 BAC transgenic mouse creation 39
2.3.2 Genetrap knockout mouse creation (chimera) 39
2.4 Mouse colony maintenance 40
2.4.1 Housing conditions 40
2.4.2 Mating 40
2.4.3 Euthanasia 41
2.5 Mouse embryo manipulation 41
2.5.1 Xgal staining 41
2.5.2 Cartilage staining 42

2.6 Mouse genotyping 43
2.6.1 Human BAC transgene detection 43
2.6.2 Mouse pGT1 genetrap detection 43
2.7 *Cecr2* knockout mice rescue experiments 44
2.7.1 Human BAC transgene rescue 44
2.7.2 Folic acid rescue 45
2.8 RNA manipulation 46
2.8.1 Total RNA isolation 46
2.8.2 RNA electrophoresis 46
2.8.3 Poly-A$^+$ (mRNA) RNA isolation 47
2.9 Probe labeling 47
2.9.1 Klenow labeling 47
2.9.2 PCR labeling 48
2.10 Hybridizations 48
2.10.1 Southern 48
2.10.1 Northern 49
2.11 Protein manipulation 49
2.11.1 Insectselect™ protein expression 49
2.11.2 Transfection of insect cell lines 50
2.11.3 Generation of stable recombinat insect cell lines 51
2.11.4 Recombinant protein isolation/purification 51
2.11.5 Protein electrophoresis 52
2.11.6 Polyacrylamide gel staining 53
2.11.7 Electroblotting 53
2.11.8 Western blotting 53
2.11.9 Histone acetyltransferase assay 54
3. RESULTS **56**
3.1 Gene mapping in the CESCR 56
3.2 Cloning/characterization of *SLC25A18* 56
3.3 Localization and Northern analysis of *ATP6E* 58
3.4 Identification of CECR2 59
3.5 Identification and cloning of a BAC 357F7 genomic sequencing gap 62
3.6 Northern analysis of *CECR2* 65
3.7 Identification of *CECR2* alternative transcripts 66
3.8 Prediction of *Cecr2* genomic structure and mRNA sequence 68
3.9 Sequence analysis of *CECR2/Cecr2* 69
3.10 Expression of CECR2 in insect cells 71
3.11 Histone acetyltransferase assay of CECR2 72
3.12 Creation/characterization of *CECR2* transgenic mice 73
3.13 Creation of *Cecr2* 'knockout' mice 74
3.14 RT-PCR of *Cecr2* in *Cecr2$^{pGT1/pGT1}$* mice 77
3.15 Reporter gene expression of Cecr2 in mouse embryos 77
3.16 Skeletal analysis of *Cecr2* mutant embryos 79
3.17 Folic acid rescue experiment in *Cecr2* mutant mice 79
3.18 *CECR2* transgene complementation experiment 80
3.19 Excision of a P-element within *Drosophila CECR2* (*CG10115*) 82

4. DISCUSSION **117**
4.1 Identification of 3 genes in the distal CESCR 117
4.1.1 ATP6E, a subunit of vacuolar ATPase 117
4.1.2 SLC25A18, a mitochondrial glutamate carrier 118
4.1.3 CECR2, a putative transcriptional regulator 120
4.2 *CECR2* encodes a putative transcriptional regulator 120
4.3 *CECR2* has a complex expression pattern 125
4.4 *CECR2* over-expression in mouse produces conflicting results 128
4.5 Human *CECR2* containing BAC does not rescue *Cecr2* mutants 131
4.6 *Cecr2* is potentially important for neurulation 132
4.7 Folic acid cannot rescue *Cecr2* mutants 136
4.8 Cat eye syndrome: monogenic or polygenic? 139
4.9 Future directions 141
4.9.1 Characterization of the CECR2/Cecr2 protein 141
i) Interacting proteins 141
ii) DNA binding experiments 142
4.9.2 Further characterization of Cecr2 mutants 142
i) Generation of a Cecr2 null allele 142
ii) Tunel staining to observe apoptosis patterns 143
iii) Folic acid rescue experiment 144
iv) CECR2 transgene rescue experiment 145
v) Comparative gene expression between Cecr2 genotypes 146
vi) Generation of double mutants with other NTD genes 146
4.9.3 Generation of a Drosophila mutant 147
4.10 Concluding remarks 148
REFERENCES **150**
APPENDIX **178**

LIST OF FIGURES

Figure 1. Chromosome 22 genomic rearrangements....3
Figure 2. Gene positions in the CESCR and the mouse orthologous region....33
Figure 3. Gene content of human BACs 357F7 and 77H2....88
Figure 4. Expression profile of *SLC25A18*....89
Figure 5. Expression profile of *ATP6E*....90
Figure 6. *CECR2* genomic structure, primer and EST locations....91
Figure 7. Genomic 'RACE' of region surrounding *CECR2* exon 1....92
Figure 8. BAC 357F7 Southern blots....93
Figure 9. Schematic of region surrounding the BAC 357F7 'deletion'....94
Figure 10. Expression profile of *CECR2*....95
Figure 11. *CECR2* alternative transcripts....96
Figure 12. Comparison between the CECR2 and Cecr2 amino acid sequences....97
Figure 13. Location of CECR2 motifs....98
Figure 14. Locations of sequence homology between 3' UTRs of *CECR2* and *Cecr2* ...99
Figure 15. Western blot of transiently expressed Insectselect™ proteins....100
Figure 16. Western blot of cell line expressed Insectselect™ proteins....101
Figure 17. Southern blot and PCR genotyping of *CECR2* transgenic mice....102
Figure 18. Overview of mouse genetic manipulations....103
Figure 19. *Cecr2* knockout chimeric mouse and progeny....104
Figure 20. Organization of *Cecr2* genetrap allele....105
Figure 21. Phenotypes of *Cecr2*$^{pGT1/pGT1}$ mice....106
Figure 22. RT-PCR of *Cecr2* in *Cecr2*$^{pGT1/pGT1}$ embryos....107
Figure 23. *Cecr2* reporter gene expression in 13.5 dpc mouse embryos....108
Figure 24. *Cecr2* reporter gene expression in 9.5 dpc mouse embryos....109
Figure 25. *Cecr2* reporter gene expression in 10.5 dpc mouse embryos....110
Figure 26. *Cecr2* reporter gene expression in 13.5 dpc cleared mouse embryos....111
Figure 27. Sections of *Cecr2* reporter gene expression in 13.5 dpc mouse embryos....112
Figure 28. Skeletal analysis of *Cecr2* mutant mice....113
Figure 29. *CECR2* transgene experiment mating scheme....114
Figure 30. *Drosophila CG10115* intron 1 characterization....115
Figure 31. *Drosophila P*-element mobilization mating scheme....116
Figure A1. Map of pGEM-Teasy vector....184
Figure A2. Map of pBS-II-SK^{+} vector....185
Figure A3. Location of *CECR2* expression constructs....186
Figure A4. Map of Insectselect™ vectors pMIB/V5-His A-C....187
Figure A5. Polylinker of modified pMIB/V5-His A vector....188
Figure A6. 5' RACE methodology....189
Figure A7. 'Genomic' RACE methodology....190

LIST OF TABLES

Table 1. Repeatmasker summary of CESCR region 10
Table 2. Comparison of exon/intron sizes between *CECR* and *Cecr2* 84
Table 3. Genotype/phenotype data of (129P2:BALB/c) *Cecr2* genetrap mice 85
Table 4. Results of maternal folic acid supplementation experiment 86
Table 5. Results of *CECR2* transgene rescue experiment 87
Table A1. Primers for human genes discussed in this study 179
Table A2. Primers for mouse genes discussed in this study 182
Table A3. Primers for vectors used in this study 183
Table A4. Primers for *Drosophila* genes discussed in this study 183

LIST OF ABBREVIATIONS

aa.....amino acids
BAC.....bacterial artificial chromosome
BLAST.....basic local alignment search tool
bp.....base pairs
BSA.....bovine serum albumin
CBP.....CREB binding protein
CES.....cat eye syndrome
CESCR.....cat eye syndrome critical region
Ci.....cubitis interruptis
CNS.....central nervous system
Der.....derivative
DGS.....DiGeorge syndrome
DNA.....deoxyribonucleic acid
dpc.....days post coitum
Dup.....duplication
EDTA.....ethylene diamine tetraacetic acid disodium salt
ER.....estrogen receptor
ES cell.....embryonic stem cell
EST.....expressed sequence tag
HAT.....histone acetyltransferase
HSA.....Homo sapiens
HSLAS.....Health Services Lab Animal Services
Inv.....inverted
kbp.....kilobase pairs
kDa.....kilodaltons
LB.....Luria broth
LCR.....low copy repeat
LINE.....long interspersed nuclear element
M.....molar

Mbp.....megabase pairs
MIM.....Mendelian inheritance in man
mg.....milligrams
mL.....millilitres
MMU.....Mus musculus
mol.....moles
mRNA.....messenger RNA
MOPS.....3-[N-Morpholino]propanesulfonic acid
N-CoR.....nuclear receptor co-repressor
NLS.....nuclear localization signal
nt.....nucleotides
ORF.....open reading frame
P/CAF.....p300/CBP associated factor
P/CIP.....p300/CBP interacting protein
PCR.....polymerase chain reaction
pmol.....picomoles
RACE.....rapid amplification of cDNA ends
RAR.....retinoic acid receptor
RIP-140.....receptor interacting protein-140
RNA.....ribonucleic acid
RT.....room temperature
RT-PCR.....reverse transcriptase PCR
RTS.....Rubinstein-Taybi syndrome
SDS.....sodium dodecyl sulfate
SRC-1.....steroid receptor co-activator 1
TAPVR.....total anomalous pulmonary venous return
TBS.....Townes-Brocks syndrome
TOF.....Tetrology of Fallot
TR.....thyroid hormone receptor
Tris.....(hydroxymethyl) methylamine
U.....units

UTR.....untranslated region

VCFS.....Velo-Cardio Facial syndrome

v/v.....volume per volume

w/v.....weight per volume

YAC.....yeast artificial chromosome

1. INTRODUCTION

1.1 Gene dosage

Most genes are present in two copies in the human genome. Of these genes, a subset will be sensitive to abnormal gene dosage. Due to the dosage sensitivity of many genes, pressure is strong to maintain the appropriate copy number. For this reason, large-scale duplications and deletions are uncommon in liveborns as compared to single base pair mutations. Duplications on the order of an entire chromosome create severe developmental abnormalities, and embryos bearing these extra chromosomes are usually quickly aborted or fail to implant. Only 3 of the 22 human autosomes allow the fetus to survive to birth in a trisomic state with any regularity (13, 18, and 21), and often these fetuses are only mosaic for the extra chromosome (Nussbaum *et al.* 2001). Since altered gene dosage leads to abnormal development, many researchers have attempted to understand more about normal development via the study of cases of abnormal development, precipitated by altered dosage. One particular example of a chromosome in which various forms of abnormal dosage have been identified is chromosome 22

1.2 Chromosome 22

The most common deletion present in human is the 22q11 deletion, with an estimated frequency of 1 in 4000 live births (Burn and Goodship 1996). Chromosome 22 can also harbour large-scale duplications, from the order of several million base pairs (Mbp) in cat eye syndrome, to the entire chromosome in rare cases of trisomy 22. Because the chromosome is small (~50 Mbp) and gene rich, it has been intensely studied and was the first human chromosome to be sequenced (Dunham *et al.* 1999). Of the many

disorders mapping to chromosome 22, 5 are notably associated with altered gene dosage: Di-George syndrome (DGS; MIM 601362), Velo-Cardio Facial syndrome (VCFS; MIM 192430), Cat eye syndrome (CES; MIM 115470), Der(22) syndrome, and the 22q13.3 deletion syndrome (MIM 606232).

The 22q13.3 deletion syndrome involves terminal deletions of varying sizes of the long arm of one copy of chromosome 22 (Nesslinger *et al.* 1993), and presumably leads to haploinsufficiency for one or more genes in the region. The resulting phenotype is predominantly neurological in nature (Nesslinger *et al.* 1993; Phelan *et al.* 2001). The remaining four syndromes involve rearrangements of 22q11.2 (**Figure 1**). Di-George syndrome (DGS) and Velo-Cardio Facial syndrome (VCFS) are collectively known as 22q11 deletion syndrome as they correspond to the same deletion. The 22q11 deletion syndrome results from inappropriate recombination between low copy repeats (LCR) spanning the deleted region (Halford et. al 1993; Edelmann *et al.* 1999). There are 3 LCR modules in 22q11.2 that facilitate two deletions, one of 1.5 Mbp and the other 3.0 Mbp (Endelmann *et al.* 1999), both of which lead to DGS/VCFS. Notable phenotypes include conotruncal heart defects, characteristic facies, pharyngeal arch defects and psychiatric disorders (Shprintzen *et al.* 1978).

Cat eye syndrome (CES) results from a duplication, or triplication, of 22q11 (Schinzel *et al.* 1981; Reiss *et al.* 1985; McDermid *et al.* 1986; Knoll *et al.* 1995), a region centromeric to the DGS/VCFS deletion region. Typically, this takes the form of a supernumerary chromosome carrying an inverted duplication of the entire p arm and the proximal region of the q arm.

Der(22) syndrome results from a constitutional t(11;22) translocation followed by meiotic non-disjunction leading to progeny that posses three copies of 22q11 and three copies of 11q (Zakai and Emanuel 1980). This syndrome has significant phenotypic overlap with Cat eye syndrome as would be expected due to the region duplicated, yet there are differences as well (Fraccaro *et al.* 1980; Van Hove *et al.* 1992). The translocation breakpoint appears to occur in the second of the three LCR modules in the DGS/VCFS region of chromosome 22 (Funke et. al 1999; Shaikh *et al.* 1999a).

Our lab was involved in the cloning of an approximately 2 Mbp region of 22q11.2 centromeric to the DGS region. This region had earlier been linked to cat eye syndrome (McDermid *et al.* 1986; Mears *et al.* 1995).

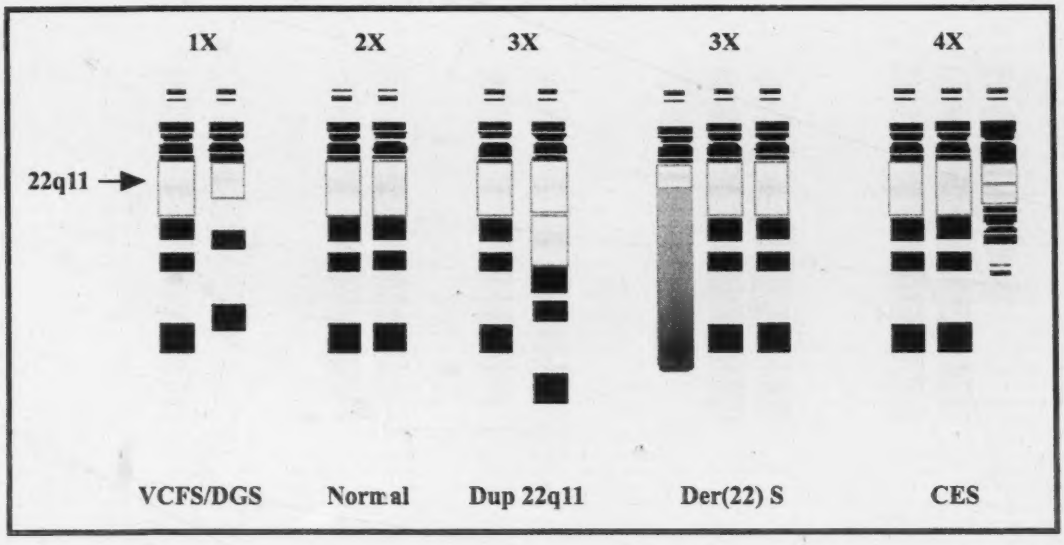


Figure 1. Representation of the different chromosomal rearrangements involving chromosome 22q11 (shown in blue), and the resulting copy number of 22q11 (image modified from McDermid and Morrow 2002).

1.3 Cat eye syndrome (CES)

1.3.1 CES history

Cat eye syndrome (MIM 115479) is associated with a gain of genetic material derived from chromosome 22q11. It is a rare (1:50000 – 1:150000) syndrome resulting in a highly pleiotrophic phenotype (Schinzel *et al.* 1981). Features may include some or all of the following: mono or bilateral iris coloboma, anal atresia with or without fistula, down-slanting palpebral features, preauricular tags and/or pits, hypoplastic kidney and congenital heart defects (Tetrology of Fallot (TOF), or Total anomalous pulmonary venous return (TAPVR)) in addition to other less common features.

These features are typically the result of a chromosome 22 derived supernumerary chromosome, which is bisatellited, isodicentric and contains an inverted duplication (inv dup(22)(q11)). This supernumerary chromosome contains two copies of 22q11, in addition to the two normal chromosome 22s, resulting in 4 copies of the region (Schinzel *et al.* 1981; McDermid *et al.* 1986).

CES marker chromosome breakpoints fall within the same LCR regions that are found in the DGS region of 22q11.2, suggesting a common mechanism for the rearrangements (McTaggart *et al.* 1998). CES marker chromosomes are commonly found in two sizes; the difference is apparently associated with which breakpoint is used in the DGSCR (McTaggart *et al.* 1998). This finding also shows that the DGS region is duplicated in CES patients with no apparent phenotypic outcome, based on comparisons between CES patients with different size duplications (see *1.3.4*). However, other cases of CES have been reported with no observable marker chromosome (Reiss *et al.* 1985; Knoll *et al.* 1995). In these two cases, a visible interstitial duplication of 22q11 results in

3 copies of the region (**Figure 1**). This would suggest that three copies of 22q11 are sufficient to produce the CES phenotype.

1.3.2 Disorders that share phenotypic overlap with CES

As there are several disorders with phenotypes overlapping with various degrees to CES, their root causes are of interest with respect to CES. We hope to learn more about CES if we understand more about other causes of the abnormalities found in CES patients. While pathways leading to particular abnormalities may not be shared between CES and other disorders, the understanding of these pathways may help us formulate better hypotheses about CES.

One syndrome with CES-like features is der(22) syndrome. der(22) syndrome results from meiotic non-disjunction events in carriers of a balanced translocation between chromosomes 11 and 22, leading to extra copies of 22q11 and 11q23 on the supernumerary derivative chromosome (Zackai and Emanuel 1980; Schinzel *et al.* 1981b). der(22) syndrome patients show karyotypes of 47,XX (or XY), +der(22), t(11;22)(q11;q23). There is significant overlap between the phenotypes of der(22) syndrome and CES (Fraccaro *et al.* 1980), but also pronounced differences. Coloboma and TAPVR are not seen in der(22) patients. A possible reason for this is the presence of modifier loci on chromosome 11, which may suppress the aforementioned phenotypes when present in three copies. The region of chromosome 22 that undergoes the recombination with chromosome 11 is found within one of the LCR elements within the DGSCR (Shaikh *et al.* 1999b).

Another disorder not related to chromosome 22 with phenotypic overlap with CES is Townes-Brocks syndrome (TBS, MIM 107480), an autosomal dominant disorder that presents with a highly variable phenotype including imperforate anus, preauricular tags or pits, anomalies of the hands and feet and renal abnormalities (reviewed in Powell and Michaelis 1999). It has been shown that mutations in the gene encoding the repressor SALL1 lead to TBS (Kohlhase *et al.* 1998) and that SALL1 binds to pericentromere heterochromatic regions of chromosomes, and possesses intrinsic transcriptional repression abilities (Netzer *et al.* 2001). This is of interest in light of the fact that the CESCR is directly adjacent to the chromosome 22 pericentromere.

Renal coloboma syndrome (RCS, MIM 120330), also known as Papillorenal syndrome, typically presents with ocular coloboma and renal hypoplasia or various other kidney defects with some overlap to CES (reviewed in Eccles and Schimmenti 1999). Mutations in the paired box transcription factor *PAX2* have been implicated in RCS (Sanyanusin *et al.* 1995; Schimmenti *et al.* 1995). Mice with homozygous *Pax2* mutations develop iris coloboma (Torres *et al.* 1996), in addition to ear and kidney defects (Torres *et al.* 1995; Favor *et al.* 1996). Haploinsufficiency of *Pax2* during mouse development leads to exencephaly (Torres *et al.* 1996), while over-expression of *Pax2* leads to lethality in *Drosophila* (Kavaler *et al.* 1999) or kidney defects in mice (Dressler *et al.* 1993), suggesting that developmental dosage of this gene is critical.

CHARGE association (MIM 214800) bears striking similarity to the CES phenotype. Features include **c**oloboma, **h**eart defects, choanal **a**tresia, mental **r**etardation, **g**enital hypoplasia and **e**ar anomalies (Pagon *et al.* 1981; Blake *et al.* 1998). However, CHARGE patients do not bear a supernumerary marker chromosome, nor has strong

linkage been observed to one particular locus in the genome, suggesting complex inheritance or *de novo* mutations in each patient. In addition, *PAX2* has been tested in 34 CHARGE patients and no mutations were observed, excluding it from the CHARGE association (Tellier *et al.* 2000). However downstream targets of the transcription factor PAX2 would remain as candidate genes in CHARGE. It would also be of interest to determine if any of the genes in the CESCR are potentially regulated by PAX2, which would lend credence to the hypothesis that CES and RCS share some developmental etiology.

1.3.3 Monogenic vs polygenic

Both Townes-Brock syndrome (TBS) and Renal Coloboma syndrome (RCS) display autosomal dominant inheritance and as such are monogenic in nature. TBS and RCS are caused by mutations in genes that are directly involved with gene expression; *PAX2* being a transcription factor proper and *SALL1* being a transcriptional repressor possibly modifying chromatin structure (Netzer *et al.* 2001). TBS and RCS both have a pleiotrophic phenotype with significant overlap with CES. This begs the question: is CES a single gene disorder? Could the over-expression of a single gene lead to a phenotype similar to TBS and RCS, which are caused by loss of function mutations? If one were to envision a gene which regulates CES genes, then perhaps duplication of the CES genes could sequester more of the upstream regulatory protein, leading to a relative shortage of that factor. Alternatively, if the upstream protein controlling CES gene expression is limiting, having four copies of the CES region may alter regulation of the CES gene(s), allowing them to be expressed in a temporally or spatially inappropriate manner. This

could be tested by measuring expression levels and patterns of CES genes in a mouse over- or under-expressing *PAX2* (Dressler *et al.* 1993) or *SALL1*.

1.3.4 Delineating the CES critical region (CESCR)

Much of the initial work on CES in the McDermid lab involved delineating the CESCR and creating a physical map of 22q11. Through the use of several patients with unique duplications, dosage analysis with probes throughout 22q11 narrowed the CESCR (Mears *et al.* 1994). It was also determined that the CESCR and DGSCR were mutually exclusive. Further refinement based on a patient with supernumerary minute double ring chromosome 22 narrowed the CESCR to a 2 Mbp region between the centromere and the gene *ATP6E* (Mears *et al.* 1995), the first gene mapped to the CESCR (Baud *et al.* 1994). The duplication breakpoints of a patient with an interstitial duplication of 22q11.2 allowed the proximal boundary for the CESCR, approximately 1 Mbp from the centromere, to be assigned (H. McDermid, unpublished data). A caveat in this delineation of the critical region was that the interstitial duplication patient did not have all the CES features, suggesting that either genes for the missing phenotypes were not duplicated or these phenotypes were non-penetrant in this individual. This delineation of the CESCR subsequently allowed a physical map of the region to be made. Both a YAC (McDermid *et al.* 1996) and a BAC/PAC (Johnson *et al.* 1999) contig were created, and the latter was subsequently sequenced as part of the chromosome 22 sequencing project (Dunham *et al.* 1999) and ultimately the human genome project (International Human Genome Sequencing Consortium, 2001).

1.3.5 Sequence analysis of the CESCR

A portion of the CESCR sequence analysis is contained within this thesis and will be discussed later. Much of what follows in this section is described in detail in Footz *et al.* 2001. The sequences of CESCR in human and mouse region of conserved synteny were analysed by computational methods to identify putative genes and conserved elements within the 1.5 Mbp human and 450 kbp mouse contigs (Footz *et al.* 2001). Based on gene prediction, EST mapping and comparative homology with the mouse, 14 putative genes were identified in human, 10 of which had putative orthologs in mouse (**Figure 2**). Genes were first categorized into previously characterized or uncharacterized genes. The uncharacterized, putative genes were given generic names based on their relative position within the CESCR (*CECR1-CECR9*). Previously identified genes included *IL-17R*, *APT6E*, *MIL1* and *BID*. One additional gene was uncharacterized when identified, but based on its sequence (see section **3.2**) was placed into the mitochondrial solute carrier gene family and was given the name *SLC25A18* by the human gene nomenclature committee.

Sequence of the human contig revealed an interesting contrast between the proximal 400 kbp and distal 700 kbp with respect to sequence composition. The distal 700 kbp between the genes *IL-17R* and *BID* showed conservation of synteny with the mouse contig which mapped to 62-63 cM from the centromere of mouse chromosome 6 (Footz *et al.* 1998). The proximal region however, is not conserved in the linkage group on mouse chromosome 6 and appears to be specific to the primate lineage (see below). There was a marked difference in G+C content and repeat distribution; the proximal region being lower in % G+C and LINE-rich elements, while the distal region was higher

in % G+C and LINE-poor while SINE-rich (**Table 1**) (Footz *et al.* 2001). The distal region of the CESCR contained sequence unique to chromosome 22 while the proximal

Table 1. Summary of Repeatmasker output analysing sequence features of the 1.1 Mbp CESCR (from Footz *et al.* 2001).

	Proximal of *IL-17R*	***IL-17R* and Distal**
Length (bp)	402,165	678,708
% G+C	41.1	46.4
% SINES	10.5	36.1
% LINES	29.3	9.2
% LTR elements	8.5	2.80
% DNA elements	1.4	2.5
% Unclassified repeats	0.2	0.0
% Total repeats	50.0	50.6

region was comprised of a patchwork of duplicated sequences from various regions of the genome. Concordant with the sequence differences (**Table 1**) between the proximal and distal CESCR, the novel genes were clustered in the distal 700 kbp, whereas only 2 genes were observed in the proximal 400 kbp. These putative genes (*CECR7* and *CECR8*) are not present in mouse and appear to be comprised of duplicated portions from elsewhere in the genome, created through an exon shuffling mechanism (Bridgland *et al.* 2003). It does not appear from the sequence that *CECR7* and *CECR8* encode functional proteins, although this conclusion cannot be validated without further experimentation.

Thus, the efforts in elucidating candidate genes for CES have focused on the 700 kbp of unique sequence between the genes *IL-17R* and *BID*. Of the 12 genes in this distal region, 10 have putative orthologs on mouse chromosome 6. *CECR1* and *CECR4* do not

appear to be present within the mouse contig and *CECR1* may in-fact be entirely absent from the mouse genome (S. Maier, unpublished data).

1.3.6 Candidate gene approach

Of the genes identified within the CESCR, a hierarchy was created with respect to their likelihood as candidates. This was based on a combination of the pattern of gene expression, postulated gene function if known, and whether said function was likely to be dosage sensitive. Genes that have been shown to be dosage sensitive are commonly found in intermolecular complexes with defined stoichiometry (Fisher and Scambler 1994). Haploinsufficient genes fall into categories such as transcription factors, transcriptional regulators, growth factors, structural proteins and receptors. The diseases caused by haploinsufficient genes are dominant and examples include collagen (osteogenesis imperfecta; Byers 1993), elastin (Williams syndrome; Mari *et al.* 1995), peripheral myelin protein 22 (Hereditary neuropathy with pressure palsies; Chance *et al.* 1993) and the *GLI-3* transcription factor (Grieg cephalopolysyndactyly syndrome; Kalffe-Suske *et al.* 1999). Keeping in mind that there are no absolutes in these categorizations, these criteria allow a starting point for allocation of lab resources into the study of the 14 CESCR genes.

Previously, *CECR1* expression has been shown in several tissues affected in CES, notably the outflow tract of the heart (Riazi *et al.* 2000) making it an attractive candidate for CES heart defects. In addition, the protein sequence is similar to Adenosine Deaminase (ADA), several invertebrate growth factors including a family of 6 *Drosophila* growth factors, some of which have been shown to have mitogenic growth

factor activity via their intrinsic ADA activity (Zurovec *et al.* 2002, Maier *et al.* 2001). Proteins of these classes may be sensitive to dosage and, as such, *CECR1* will be investigated further. ADA activity has not yet been shown for *CECR1*, but this may be due to technical difficulties (H McDermid, personal communication). *CECR1* is considered a viable candidate for CES, though the lack of a murine ortholog precludes mouse modeling of CES with a *CECR1* transgene.

Of the remaining genes in the CESCR, two have been previously characterized: *IL-17R* and *ATP6E*. *IL-17R* is a receptor for the T-cell cytokine IL-17, which is a small, secreted molecule involved in the inflammatory response (Yao *et al.* 1995). While *IL-17R* is ubiquitously expressed (Yao *et al.* 1997), its ligand, IL-17 is only expressed in T cells (Yao *et al.* 1995) and as such over-expression of the receptor would seem to be inconsequential. Considering that there are no T-cell defects in CES, *IL-17R* was given a low priority status. *ATP6E* is the epsilon subunit of vacuolar ATPase (V1), a protein involved in acidification of vacuoles and other intracellular organelles. The epsilon subunit is a regulatory subunit of the v-ATPase and is part of a multi-subunit protein molecule. While it could theoretically be dosage sensitive, we chose to place it lower on the list for study based on studies of other V-type ATPases. For example, mutations in *ATP6V0A4* (Smith *et al.* 2000) and *ATP6V1B1* (Karet *et al.* 1999) cause autosomal recessive renal tubular acidosis.

The remaining genes in the CESCR have been previously described (Footz *et al.* 2001), but will be mentioned briefly here. A CESCR gene placed lower in the hierarchy was *CECR5*. It encodes a protein with significant similarity to the enzyme CDP-alcohol phosphatidyltransferase from the fission yeast *S. pombe*. Such enzymes are involved in

fatty acid metabolism (Antonsson 1997) and as such we did not feel it was an attractive candidate for such defects as congenital heart defects or kidney hypoplasia. *CECR3* is conserved in mouse but does not appear to encode a protein. Potentially a functional untranslated RNA, this makes further study of *CECR3* quite challenging and therefore, this gene was not given high priority. *CECR9* was identified based on a trapped exon and homology to the mouse. There are no ESTs for this transcript, and based on Northern analysis the transcript is very weakly expressed, with the exception of heart, where significant expression was observed. The complete gene structure has yet to be determined for this gene. *CECR4* was identified through a trapped exon. Interestingly, this transcript partially overlaps with the 5' end of *CECR5* in the inverse orientation, suggesting a possible antisense regulation. *CECR4* is not conserved in mouse, and no significant open reading frame is predicted from the partial cDNA sequence, so it is possible that *CECR4* is a spurious transcript.

CECR6 encodes a single exon gene with a leucine zipper motif, which is commonly found in transcription factors and is involved in protein dimerization. This gene is being studied further. *CECR2* is the focus of this thesis and will be described in detail. It encodes a putative transcriptional co-activator (see **1.4** for more detail). Thus, of the 14 putative genes in the region, we have given three genes a high priority for study with respect to potential involvement in CES. In order to test if *CECR1*, *CECR2*, or *CECR6* are dosage sensitive we have generated transgenic mice, which carry extra copies of the human genes in order to determine if abnormal development results.

1.3.7 Mouse modeling of CES

Two approaches were considered in an attempt to model CES in the mouse. The first method was to engineer a duplication of the entire region in mouse through cre-lox recombination (reviewed in Misra and Duncan 2002; Tronche *et al.* 2002). The advantages to this method are that multiple genes can be duplicated at the same time and in the correct orientation with respect to each other. In addition, insertional mutagenesis, like that with BAC transgenics, is not a problem. In addition there is a known copy number on the modified chromosome. This approach would more closely mimic the actual copy number in human CES patients: 4 copies when the engineered chromosome is homozygous. Limitations with this method include the fact that targeting constructs must be created to include the appropriate recombination signals and selectable markers, all of which must be performed in ES cells (reviewed in Tronche *et al.* 2002). The second limitation is the fact that not all the genes are present within the orthologous mouse region, most notably *CECR1* (Footz *et al.* 2001). Benefits of this approach include the fact that mouse genes are being duplicated in the mouse as opposed to human genes, which are not guaranteed to function within the mouse cell. In addition, the possibility that CES may be a multigenic disorder might make it more likely to observe unusual phenotype only when multiple genes are duplicated concurrently.

The second approach to mouse modeling of CES involves transgenesis using single BAC or PAC clones that contain human genomic DNA and one or more genes in their entirety, along with possible regulatory elements (reviewed in Williams and Wagner 2000). Technically, this is much easier to perform than cre-lox mediated recombination for several reasons. Firstly, naked DNA is injected into fertilized mouse oocytes as

opposed to ES cells. No targeting construct must be built and ES cell injection into mouse blastocysts is not required. BAC transgenics are quicker and cheaper to create, but drawbacks exist. Firstly, the DNA insertion is not targeted allowing insertional mutagenesis effects. Secondly, multiple insertions are possible within the same genome and determining the insertion sites can be technically difficult. Thirdly, upon integration, BAC DNA can recombine with itself creating long arrays in a head to tail manner. This can greatly increase the copy number of the transgene, but at the same time not all copies of the transgene will be active (Dorer 1997). While absolute copy number may be ascertained, the relative copy number (based on expression) is difficult to determine and quantitative RNA or protein measurements are required to determine the actual increase in gene expression. Another concern with BAC transgenics involves potential regulatory elements up or downstream of the structural gene. These are difficult to identify, and as such, may be absent when clones are chosen for transgenesis. The last major problem involves interspecies gene function. Expressing human genes in a species diverged for tens of millions of years can be inherently problematic insofar as negative results are difficult to interpret. Is a human gene not functional in the mouse because other necessary proteins are too diverged to recognize the transgenic protein? Or has the human protein evolved a new function with respect to its mouse counterpart? It is difficult to elucidate the answers to these questions with negative results of human proteins expressed in mouse. However given the high degree of similarity between the CECR2 and Cecr2 (see **3.9**) it seems plausible that they may function interchangeably.

Considering the pros and cons of each system, technical and monetary considerations, we decided to take the candidate gene approach and express the human

candidate genes via BAC transgenesis in the mouse. The following questions exist with respect to modeling CES in mice: are mice developmentally close enough to humans that CES could be caused by over-expression of one or more genes? Will the human genes be expressed in an appropriate temporal and spatial specific manner similar to the expression of the mouse ortholog? And if not, can we extrapolate the data from ectopic expression to over-expression in native tissues? In addition, if complex gene regulation occurs in human, such as alternative promoters, alternative polyadenylation, RNA localization and degradation, and post-translational modifications, are they faithfully reproduced in the mouse? Is the mRNA even translated into a protein? These questions must be addressed in order to determine the true effects of a transgene in a complex system such as a mouse. It should be noted however, that there are many examples of transgenic mice in which a transgene correctly mirrors the human disease phenotype when over-expressed. Over-expression of large regions of human chromosome 21 which contain the *DYRK1A* (*minibrain*) gene result in cognitive defects, neurodevelopmental delay and motor abnormalities (Smith *et al.* 1997), which mimic conditions seen in trisomy 21 (Down syndrome, MIM 190685), the most common duplication in humans. These findings are confirmed when the native mouse copy of *Dyrk1a* is over-expressed (Altafaj *et al.* 2001), suggesting that the over-expression of human genes in mouse can accurately recapitulate human genetic disease in these cases.

A second example of successful transgenic modeling is Charcot-Marie-Tooth (CMT, MIM 118220), a disease caused by a duplication surrounding the *PMP22* gene on human chromosome 17. CMT is a neuropathy, which is characterized by progressive muscle weakness due to demyelination (Dyck *et al.* 1993). A successful mouse model

was generated by pronuclear injection of a human YAC containing the *PMP22* gene (Huxley *et al.* 1996), suggesting that overexpressing *PMP22* in mouse can recapitulate human disease. Interestingly, mutations in *PMP22* cause hereditary neuropathy with pressure palsies (HNPP, MIM 162500) and are mimicked by the naturally occurring Trembler mutants in mouse (Suter *et al.* 1992a; Suter *et al.* 1992b) showing that some genes are sensitive to dosage in both directions, with distinct phenotypes in each case. Taking the above examples of successful transgenic animals, we proceeded knowing that caveats exist and results may be difficult to interpret.

1.4 Transcriptional Co-activators

As *CECR2* appears to fall into the category of transcriptional co-activator (based on sequence analysis; see section **3.9**), some background information on co-activators will be presented.

The basal transcription machinery does not work in isolation; accessory factors are needed for active and efficient transcription. Transcription factors are one such group. They are prototypically DNA binding proteins with sequence specific recognition sites in DNA, which allow target specificity. Many of the homeotic genes are transcription factors. However, it has become apparent in the past 15 years that many other factors are involved in bringing a gene into a transcribable state. There appears to be a hierarchy of proteins much like an inverted pyramid that gets larger the further away from the DNA you go. TATA box promoters are bound by the protein complex TFIID, which is comprised of TATA binding protein (TBP) and various TBP associated factors (TAFs) (Dynlacht *et al.* 1991; Tanese *et al.* 1991). However, it was observed that while TBP

alone could activate transcription at basal levels, it could not increase transcription in response to enhancer-binding proteins (Pugh and Tjian 1990). This led to the hypothesis that the TAFs were in-fact co-activators, that somehow transmitted a signal to the basal transcription machinery to up-regulate transcription. It later became apparent that other complexes can interact with different transcriptional activators, such as TFIID, and that these co-regulator complexes can have distinct components, often in a tissue specific manner.

The co-regulator class of proteins can be subdivided further into co-activators, co-repressors, factors that interact with the basal machinery and others that interact with co-activators or repressors. This group of proteins is quite diverse with distinct activities. Co-regulators can be grouped into five categories based on their general properties (Lemon and Tjian 2000 and references therein): (1) factors that interact with core transcription machinery components; (2) factors that preferentially associate with co-activators or repressors; (3) multi-subunit co-activators; (4) factors that covalently modify nucleosomes; and (5) factors that utilize ATP hydrolysis in order to modify chromatin structure. Group 1 co-regulators can modulate DNA binding or modify other co-regulators, interact directly with RNA Pol II, or recognize promoter elements. Groups 4 and 5 involve chromatin remodeling. These two categories have ATP-dependent chromatin remodeling (SWI/SNF complex) or intrinsic acetyltransferase/deacetylase activities (CBP, GCN5, HDAC-1 and others). Group 4 proteins can acetylate histones, thus modifying their grip on DNA, or can activate core transcription machinery via acetylation. The above categories are broad and include many members, which typically form large complexes. It is also important to note that different transcription factors

require different co-activator complexes to effectively induce transcription, but many of these co-activator complexes contain shared components such as CREB binding protein (CBP)/p300 and p300/CBP associated factor (P/CAF) (Korzus *et al.* 1998).

1.4.1 Histone modification and gene expression

A reccurring theme in gene regulation is chromatin modification. It has long been known that histones are subject to epigenetic modification, which is somehow related to gene expression (Allfrey *et al.* 1964) or repression. It was later shown that *Tetrahymena* histone acetyltransferase A (HAT A) was related to the yeast transcriptional activator protein GCN5, which allowed the inference to be drawn that HAT activity in transcriptional regulatory proteins could influence transcription by modifying chromatin conformation (Brownell *et al.* 1996). Since then, many transcriptional co-activators have been shown to have HAT activity and it seems to be crucial for their activator capabilities.

HATs fall into four loose categories (Grant 2001 and references therein): (1) GNAT family (GCN5-related N-acetyltransferases), which includes proteins involved in transcription initiation (GCN5, P/CAF), elongation (Elp3) and telomere silencing (Hat1); (2) CBP/p300 family, which includes related proteins implicated as co-activators for transcription factors; (3) MYST family (MOZ, Ybf2/Sas3, Sas2 and Tip60), members of which have been implicated in replication, transcription elongation and mating type switching; (4) Basal transcription factors such as TAFII250 and some of the nuclear receptor co-activators such as ACTR or SRC-1.

Acetylation of the core histones destabilizes chromatin folding, and subsequently relaxes chromatin. When chromatin is relaxed via acetylation of the N-terminal histone tails, DNA/histone conformation alters allowing transcription factor access to DNA and ultimately an increase in transcription (Lee *et al.* 1993; Tse *et al.* 1998). This correlation is widely documented and now accepted as one step in the process of gene expression.

Other histone modifications include phosphorylation and methylation. Less is known about these processes as compared to acetylation, but it does appear that both correlate with increased gene expression. Phosphorylation of serine 10 in histone H3 has been implicated with gene activation in humans (Thomson *et al.* 1999) and heat shock dependent gene activation in *Drosophila* (Nowak *et al.* 2000). Methylation of histones can also contribute to transcription activation. The histone methyltransferase CARM1 specifically methylates arginine residues in histone H3, in addition to interacting with the steroid receptor co-activator GRIP-1 (Chen *et al.* 1999a). Other lysine residues on histones H3 and H4 can be multiply methylated, and are found at transcriptionally active loci (Strahl *et al.* 1999). Yet in other cases methylation can close chromatin and heterochromatize a region. Overexpression of human SUV39H1, a homolog of *Drosophila* SU(VAR)3-9, a suppressor of position effect variegation, can induce ectopic heterochromatin (Melcher *et al.* 2000).

Histone ubiquitination has also been observed with transcription factor TAFII250 (in addition to acetyltransferase activity), and this has been implicated in gene activation of Dorsal in *Drosophila* (Pham and Sauer 2000). However, chromatin remodeling does not always correlate to gene expression. The NURD complex has been shown to to have a repressive effect on transcription by its ATP-dependent chromatin remodeling activity

and histone deacetylase activity (Xue *et al.* 1998). There are likely other as yet unidentified histone modification events involved in activation or repression of genes and opening and closing of chromatin.

The observation of these various epigenetic modifications of histones has led to the proposal of a histone modification language, termed the 'histone code' (Strahl and Allis 2000). The histone modifications likely serve two purposes: (1) to modify histone-DNA contact, altering chromatin structure and (2) act as interaction signals for other proteins, which would allow these proteins to contact the histone-DNA complex, starting a protein recruitment cascade. It is hypothesized then, that different transcriptional regulatory proteins will have evolved different motifs capable of recognizing distinct histone modifications (Strahl and Allis 2000).

1.4.2 Co-activator signature motifs

Proteins will often have amino acid motifs characteristic of the family to which they belong. Transcriptional co-activators are no different, but while the motifs found within the group do not define the group, they do help to subdivide it. Once such motif is the bromodomain. It was originally described as a 60 amino acid conserved sequence found in six protein from human, yeast and *Drosophila* (Haynes *et al.* 1992). Subsequent examination of the motif, helped via large scale sequencing, showed that the bromodomain is actually around 110 amino acids in length and is found in a large number of proteins, many of which fall into the transcriptional co-activator category (Jeanmougin *et al.* 1997). Based on the categorization of the bromodomain containing proteins it was speculated that this motif might interact with chromatin. This later proved

to be correct as the structure and ligand of the P/CAF bromodomain were identified. The bromodomain forms a four-helix bundle, which specifically interacts with acetylated lysine residues on histone tails (Dhalluin *et al.* 1999). Interestingly, many of the proteins with bromodomains (eg CBP/p300 (Ogryzko *et al.* 1996), P/CAF (Yang *et al.* 1996), GCN5 (Brownell *et al.* 1996)) also possess HAT activity. This places the ability to acetylate histones and subsequently bind acetylated histones within the same molecule. This may help to tether co-activators to the DNA allowing assembly of a protein complex, although it is not clear whether these co-activators provide the initial acetylation event or bind first and acetylate after binding.

There are at least two instances of HATs being regulated through acetylation. The HAT and co-activator ACTR can be acetylated by another HAT and co-activator CBP/p300, resulting in disruption of a co-activator complex assembled around a promoter bound estrogen receptor (ER) (Chen *et al.* 1999b). In addition, the HAT and co-activator P/CAF is also capable of autoacetylation (Herrera *et al.* 1997), although it is not clear as to the function of this event. There is also an example of a HAT regulated by phosphorylation. ATF-2 is a DNA binding transcription factor that has intrinsic HAT activity, which appears to be upregulated via phosphorylation of the ATF-2 protein (Kawasaki *et al.* 2000). These post-translational modifications of HAT proteins add another level of complexity onto an already complex formula of gene regulation.

A second motif found in several co-activators is the LXXLL motif and variants thereof (Heery *et al.* 1997). This motif has been shown to be involved in protein-protein interactions between various co-activators/repressors (CBP, SRC-1, RIP-140, N-CoR) and nuclear receptors (RAR, TR, ER) and is also tied to the transcriptional activation

capabilities of these molecules (Heery *et al.* 1997; McInerney *et al.* 1998). Variants of these motifs have been identified and appear to mediate interactions with specific receptors. For example, the FXXLF motif modulates co-activator interaction with the androgen receptor (AR) (He *et al.* 2002), while I/LXXII is important for co-repressor binding to the thyroid hormone receptor (TR) (Webb *et al.* 2000). While there may be specificity in the inter-protein contacts between these motifs, they all appear to function in the assembly of larger co-activator or co-repressor complexes.

1.4.3 Mutations in selected co-activators

As transcriptional co-activators would seem to play an important role in assembly of gene activation complexes, one might assume that mutations in such genes would be detrimental. Validation of this assumption has been shown via the generation of co-activator mutations in animal models. A human disease associated with haploinsufficiency of *CBP* is Rubinstein-Taybi syndrome (RTS, MIM 180849) (Petrij *et al.* 1995). RTS is characterized by broad thumbs and (great) toes, characteristic facies, mental retardation, and in some cases heart defects. Mice with homozygous mutations in the *CBP* ortholog *Cbp*, or its paralog *p300*, display lethality between 9 and 11.5 dpc, and embryos are characterized by heart and neural tube defects (Yao *et al.* 1998). In a separate study, it was observed that mice heterozygous for *Cbp* mutations develop skeletal abnormalities similar to those seen in RTS (Tanaka *et al.* 1997). Based on this and other studies, mouse modeling has been used extensively to study human genetic disease.

Animal models have proven useful in determining genetic interactions and subsequently allowed the characterization of genetic pathways. For example, in *Drosophila* it was determined that *Cbp* is a co-activator for the transcription factor Cubitis Interruptis (*Ci*) (Akimaru *et al.* 1997a). In addition, *Drosophila Cbp* is a co-activator for the transcription factor *Dorsal* and together they regulate the expression of *Twist* (Akimaru *et al.* 1997b). The human homolog of *Ci* is *GLI3*. Haploinsufficiency of *GLI3* is known to cause Greig cephalosyndactyly syndrome, which has a partially overlapping phenotype with RTS (Vortkamp *et al.* 1991). In another striking link to the *Drosophila* pathway, haploinsufficiency of human *TWIST* is known to cause Saethre-Chotzen syndrome, the phenotype of which overlaps with RTS with respect to the facial and hand abnormalities (Howard *et al.* 1997; el Ghouzzi *et al.* 1997). Interestingly, when *Cbp* is over-expressed in *Drosophila*, it is lethal (Akimaru *et al.* 1997a), suggesting its dosage in both directions is critical. These results support the reasoning that dissecting genetic pathways in model organisms can help to elucidate human disease genes. Of particular interest to this project is the fact that mutations in *Cbp* and other co-activators lead to neural tube defects (see *1.5.2*).

1.5 Neural tube defects (NTDs)

1.5.1 Neural tube closure

Neurulation is a critical processes during embryonic development in which the neural plate grows outward into folds, which then migrate towards each other and fuse to generate a hollow cavity inside. This process happens between 8.5 and 9 dpc in mice, with closure starting at multiple sites along the embryo, although the timing and position

of fusion differs slightly between mouse strains (O'Rahilly and Muller 2002). Human neurulation starts around the 4 somite stage with fusion in the hindbrain region, with a second site of fusion occurring later in the forebrain region. (O'Rahilly and Muller 2002). The initial midbrain fusion spreads rostrally and caudally until the rostral and caudal neuropores close around the 20 somite stage. However, the process of neural tube fusion does appear to have differences between human and mouse and as such care must be used when comparing the two (Nakatsu *et al.* 2000). The cavity that is formed from neural tube closure goes on to produce the central nervous system (CNS). Perturbation of this process can lead to neural tube defects, in which portions of the neural tube do not close properly. Common neural tube defects include spina bifida, exencephaly, split face and rachischisis. Failure of closure of the caudal neuropore leads to spina bifida, while failure of closure of the rostral neuropore at the midbrain position leads to exencephaly in mice (equivalent to anencephaly in humans). Split face caused by failure of closure of the most anterior region of the rostral neuropore, while failure of closure over the entire neural tube leads to rachischisis (reviewed in Juriloff and Harris 2000). Different combinations of the above neural tube defects are also possible.

1.5.2 Mouse models of NTDs

More than 75 genes have been implicated in neurulation in mice (Colas and Schoenwolf 2001; Juriloff and Harris 2000). A significant number of the genes that cause NTDs are still of unknown biochemical function, but of those with defined function there is a diverse range. One group of NTD genes is comprised of transcription factors and transcriptional co-activators. Examples include *Pax3* (Auerbach 1954), *Cbp/p300* (Yao *et*

al. 1998), *Cited2* (Bamforth *et al.* 2001), *Cart1* (Zhao *et al.* 1996), *Twist* (Chen and Behringer 1995), *Ap2* (Schorle *et al.* 1996; Kohlbecker *et al.* 2002) and *p53* (Armstrong *et al.* 1995). There is another group of genes involved in cytoskeletal/actin regulation. This group includes *Shroom* (Hildebrand and Soriano 1999), *Mena* (Lanier *et al.* 1999), *Vcl* (Xu *et al.* 1998), *Mlp* (Wu *et al.* 1996) and *Lama5* (Miner *et al.* 1998). Folate metabolism (see section *1.5.4*) has also been implicated in NTDs as folate binding protein (*Folbp1*) mutants display exencephaly (Piedrahita *et al.* 1999).

For many mutants, death comes before mid-gestation (reviewed in Juriloff and Harris 2000) suggesting that the NTD is not the major abnormality since NTDs by themselves are not lethal *in utero*. For example *Cbp/p300* mutants are lethal by mid gestation (Yao *et al.* 1998), suggesting that they are critical proteins to various aspects of development, as would be assumed by their broad expression pattern and implication in various different cellular responses. However, if proteins were specifically involved in neurulation one might expect them to be expressed only at the appropriate time and place during development, yet this is not always the case. Some NTD proteins seem to have no abnormal phenotype other than a NTD, suggesting either functional redundancy in a non-neurulation function, or specific involvement in neurulation. Functional redundancy has also been observed in neurulation, as there have been examples of double knockouts leading to NTDs, where single knockouts alone do not. Examples of this phenomenon are the nuclear receptors *Rara/Rarg* (Lohnes *et al.* 1994) and integrins *Itga3/Itga6* (De Arcangelis *et al.* 1999). Redundancy certainly makes it more difficult to build hierarchies for NTDs, considering that many potential NTD genes may not yet have been identified due to lack of available double mutants. It is clear, however, that general trends do exist

in categorization of NTD mutants. Gene activation is clearly involved as is cellular movement requiring the cytoskeleton. How the different genes are connected to each other and to neurulation remains to be elucidated. From human and mouse studies it has been observed that genetics and environment both play a role in the etiology of NTDs.

1.5.3 Genetics, environment and dietary supplements

In contrast to mice, the genetics of NTDs in human are poorly understood. There are several reasons for this. NTDs have a relatively high frequency (~1/1000 births), yet a low recurrence rate (~2-5%) in families with at least one NTD birth (Fraser and Nora 1986). This would suggest a complex pattern of inheritance for human NTDs, as opposed to the simple Mendelian inheritance pattern of most mouse NTD mutations. Secondly, environmental influences, such as diet, vary widely between different human populations, likely adding a further level of complexity to the study of NTDs. Environmental influences must be taken into consideration as they likely play a role in the manifestation of a certain group of NTDs (see below). Thirdly, the rather heterogeneous genetic makeup of humans, as compared to laboratory mice, makes identification of NTD loci a more complicated task. Efforts to date have focused on testing human NTD samples for mutations in orthologs of mouse NTD genes. Unfortunately, the majority of samples used in such studies (Stegmann *et al.* 2001) have been spina bifida, while the majority of NTD genes tested in these studies cause exencephaly in mice. Clearly this approach is problematic. A second approach has involved aspects of maternal diet and folate metabolism.

It has been known for some time that environmental influences appear to play an important role in the development of human NTDs. In particular, maternal folate status seems to play a role (Czeizel and Dudas 1992), although this has been an area of debate (Kalter 2000). Folic acid supplementation was implemented in various countries in the 1990s as the result of studies linking maternal folate deficiency with NTDs. The benchmark study was the Medical Research Council Trial (MRC Vitamin Study Research Group 1991), which took place between 1983 and 1991 and involved 33 centers around the world. In this study, women who had previously given birth to a child with a NTD were split into different groups, one receiving 4 mg of folic acid daily, prior to and up to 12 weeks of conception, with the other groups receiving a placebo. From the study it was determined that folic acid reduced NTD recurrence rates by approximately 72% (MRC Vitamin Study Research Group 1991).

Subsequent to the linking of folate status and NTDs, several genes known to be involved in folate metabolism were tested for polymorphisms/mutations in human NTD samples. Some minor correlations have been observed between human polymorphisms and NTDs. For example embryo, but not maternal, homozygosity for a particular allele (T) of the methylenetetrahydrofolate reductase (*MTHFR*) gene was shown to increase rates of spina bifida (Shields *et al.* 1999), suggesting that the genotype/phenotype correlation only manifests during development. Another study suggested that a particular allele of the methionine synthase reductase gene (*MTRR*) can contribute to higher levels of NTDs in combination with low vitamin B12 levels (folate), or when combined with certain *MTHFR* alleles (Wilson *et al.* 1999). However, other loci not directly involved in

folate metabolism, which are clearly implicated in mouse NTDs, have not been found to correlate to human NTDs or in most cases simply have not been tested.

It has subsequently become of interest to researchers whether or not mouse NTDs can be 'rescued' by maternal supplements. Penetrance of several mouse models of NTDs can be reduced by maternal folic acid supplementation. *Cart1* (Zhao *et al.* 1996), *Crooked tail* (*Cd*) (Carter *et al.* 1999), *Splotch* (*Pax3*) (Fleming and Copp 1998) and *Cited2* (Barbera *et al.* 2002) nullizygous embryos can all be rescued to some degree by maternal folic acid administration. Yet some mutants are clearly resistant to folic acid treatment, while responding to other compounds. *Curly tail* (*Ct*) mutants are resistant to the effects of folic acid supplementation but myo-inositol is capable of some rescue (10% reduction) (Greene and Copp 1997). However, based on the mechanism of action of inositol (up-regulation of retinoic acid receptor beta via protein kinase C in the hindgut and posterior neuropore region), this rescue is likely specific to spina bifida. Other compounds have been shown to have seemingly paradoxical effects on mutant embryos; methionine, but not folic acid has also been shown to rescue NTDs in the *Axial defects* (*Axd*) mouse mutant (Essien and Wannberg 1993), whereas it exacerbated NTDs in *Splotch* mice (Fleming and Copp 1998).

Based on the seeming lack of causative mutations in human NTDs and the relative abundance of NTDs in mouse mutants, questions have arisen regarding the use of mouse as a model system for human NTDs. One must also consider the possibility that genotype/phenotype patterns are present but are simply too complex to see with current information. It is likely that multiple loci are involved in human NTDs, possibly at the same time and possibly in a heterozygous state. Heterozygosity at multiple loci has not

been investigated on a large scale in mice to date. There is, however, one example of multifactorial NTD etiology in mice. The SELH/Bc strain likely has three semi-dominant loci that contribute to its 30% rate of exencephaly (Juriloff *et al.* 2001). This mouse strain likely does not mirror the complex inheritance pattern in humans, but does offer an interesting avenue for study into the genetic interactions between the three loci. Mice and humans are developmentally similar and as such the tractability of genetically manipulatable systems such as mouse cannot be overlooked. On the other hand, the fact that human embryos appear to close their neural tubes slightly differently than mice with respect to the position of closure initiation (Nakatsu *et al.* 2000) may suggest that humans are less susceptible to NTDs than mice. For this reason multiple genes are likely involved in order for NTDs to manifest.

1.6 *Drosophila* genetic models

Drosophila has long been used as a model system due to its amenability to genetic manipulation. Being an invertebrate, *Drosophila* doesn't lend itself well to modeling NTDs, yet valuable information may still be gleaned from *Drosophila* mutants. Due to the fact that genetic screens can be performed in a system like *Drosophila*, many more options are available to the researcher in terms of discovering genetic interactions and genetic pathways. For this reason generation of mutants in *Drosophila* for human genes of interest is not uncommon. For example *CBP*, implicated in NTDs, has been studied in *Drosophila*, which has allowed pathways to be determined in humans (see *1.4.3 Mutations in selected co-activators*). A common mutagenesis technique involves the use of transposable element insertional mutagenesis (commonly P-elements). Large stock

centers collect and catalog these P-element containing strains, and they are available to researchers free of charge. In addition, the excision of P-elements can be performed to generate deletions surrounding the P-element insertion site. If an abnormal phenotype is observed due to a mutation in a particular gene, suppressors (or enhancers) of the phenotype may be found by further mutagenesis of the mutant strain. Suppression (or enhancement) of the mutant phenotype may be indicative of a physical or genetic interaction with the original mutant locus. Identification of these interacting loci could allow genetic pathways to be constructed, which would help us understand more about the process in humans and mice.

1.7 Main objectives of this work

At the time this project was started, there was no sequence available for the CESCR and the only gene mapped to the CESCR was *ATP6E* (Baud *et al.* 1994). At that time, the major efforts in our lab were on identification of genes that had the potential to be dosage sensitive. This would allow us to study abnormal development in the context of abnormal gene copy number. Specific goals of my project were:

1. Using computer aided sequence analysis to identify genes in the distal CESCR.
2. Clone any identified genes.
3. Analyze the expression profile of identified genes via RT-PCR and/or Northern analysis.
4. Perform sequence analysis of identified genes to learn more about potential function.

After *CECR2* was identified, it became the focus of the project with the following aims:

5. Express CECR2 in insect cells and test it for histone acetyltransferase activity.
6. (Over)express *CECR2* in mouse to determine if CES features result.
7. Determine if the human transgene is expressed in a tissue specific manner when present in mouse.
8. Create a 'knockout' or 'knockdown' of the *Cecr2* in mouse, with the use of an ES cell line bearing a genetrap within *Cecr2*, and characterize it with the following experiments:
    i. Breed the mutation to homozygosity and observe the abnormal phenotype, if any.
    ii. Determine the *LacZ* reporter gene expression profile of genetrap locus (*Cecr2*).
    iii. Determine whether or not $Cecr2^{pGT1/pGT1}$ mutants can be rescued by maternal folic acid supplementation.
    iv. Cross the human *CECR2* transgene into the $Cecr2^{pGT1/pGT1}$ knockout line to determine whether or not the *CECR2* transgene is making protein, and whether or not this protein can replace the missing mouse protein.

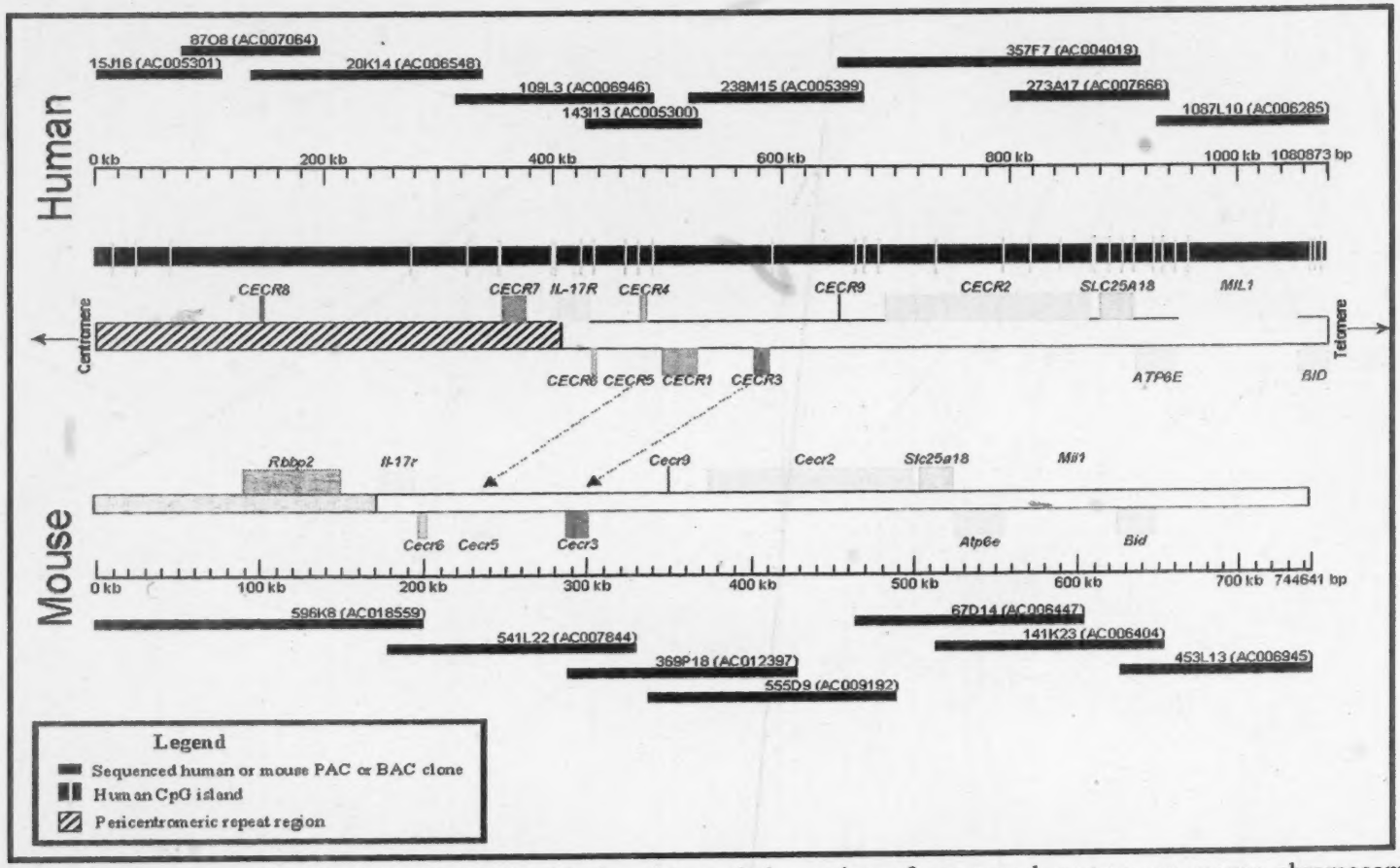


Figure 2. Putative genes identified in the CES critical region and the region of conserved synteny on mouse chromosome 6. Sequenced genomic clones are shown above and below each scale bar. Genes on top of each depicted chromosome are oriented centromere to telomere, and genes below the depicted chromosomes are oriented telomere to centromere. The grayed out section represents the portion of mouse chromosome 6 orthologous to human chromosome 12p13 (image modified from Footz et. al. 2001).

2. MATERIALS AND METHODS

2.1 DNA Isolation

2.1.1 Plasmid

Plasmid DNA was isolated by two different methods; the traditional phenol-chloroform method (Sambrook and Russell 2001), or by QIAprep spin columns as per the manufacturers instructions (Qiagen). All isolations were from 5mL LB overnight cultures, grown shaking at 37 °C in the presence of 100 μg/mL ampicillin.

2.1.2 Mouse Genomic

Mouse genomic DNA for genotyping animals was isolated from two different sources, mouse tail-clippings and extraembryonic membranes. The DNA isolation procedure for both sources was the same. Tissue samples were placed in 0.6 or 1.5 mL Eppendorf tubes in the presence of 350 μL of proteinase K buffer (50 mM Tris-HCl, 100mM EDTA, 100mM NaCl, 1% SDS) and 500 μg of proteinase K (Invitrogen) overnight at a temperature between 56 and 65 °C. The following day protein was precipitated by the addition of 125 μL of 5M NaCl followed by incubation on ice for 5 minutes. Samples were centrifuged in a microcentrifuge for 10 minutes at 14,000 rpm to pellet the precipitated proteins. The supernatant was removed to a new tube and to it was added: 1/10 volume of 3M NaOAc, 1 μL of Pellet Paint (Novagen) and 1 volume of 2-propanol to precipitate the DNA. Samples were inverted several times and spun at 14,000 rpm for 15 minutes. The supernatant was removed and the DNA pellet washed in 70% ethanol and further centrifuged at 14,000 rpm for 5 minutes. The supernatant was removed and the pellets dried before resuspending them in 20μL of dH_2O.

2.1.3 BAC DNA purification for pronuclear injection (transgenic mouse creation)

Previously isolated human BAC DNA (by Qiagen column) was further purified on a Sepharose column, for use in pronuclear injection of fertilized mouse eggs (for the creation of transgenic mice). BAC DNA was purified on a Sepharose CL-4B column (Pharmacia). The column was created in a glass chromatography column 1 cm wide by 25 cm deep. Sepharose slurry was added to the column and allowed to settle to create an 18 cm bed of sepharose. Injection buffer (10mM Tris-HCl pH 7.5, 0.1 mM EDTA) was added to the top of the column and then allowed to flow through to equilibrate the column. Once the column was full of injection buffer, 100 μL of dye (6% glycerol, 0.2% bromphenol blue) and the BAC DNA was added to the top of the column and the flow started. Immediately following DNA addition, 500 μL fractions were collected from the bottom of the column and saved for analysis on a pulsed field electrophoresis gel.

2.1.4 Agarose embedded DNA

DNA samples from agarose gels were isolated with Glassmilk (Bio101). Briefly, the agarose containing the DNA band was cut out and to it was added 3 volumes of 6M NaI. The sample was heated at >45 °C until the agarose dissolved at which time 10 μL of glassmilk was added and the sample was placed on ice for 5 minutes. The sample was then spun at 14,000 rpm in a microcentrifuge for 30 seconds to pellet the glassmilk. The supernatant was removed and the pellet washed three times in 400 μL of wash buffer (50% ethanol, 1X STE [10 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA]). The resulting pellet was resuspended in 10-15 μL of dH_2O and heated for 2 minutes at 60 °C,

followed by centrifuging for 1 minute at 14,000 rpm in a microcentrifuge. The resulting supernatant was removed and contained the DNA.

2.2 PCR and cloning

2.2.1 PCR buffer

A homemade 10X PCR buffer was used in all PCR reactions (with the exception of reactions in which high fidelity enzyme was used, in which case the manufacturer supplied buffer was used). The 10X buffer consisted of 100 mM Tris-HCl pH 9, 500 mM KCl, 15 mM $MgCl_2$ and 0.2 mg/mL BSA (Fraction V). It was used in conjunction with recombinant *Taq* DNA polymerase, (generated by Dr. Pickard, Department of Biological Sciences, University of Alberta).

2.2.2 Reverse Transcriptase PCR

RT-PCR was performed on various source RNA. For *CECR2* cloning purposes, 0.1-0.5 μg of human placental mRNA was used. For all other RT-PCR 2-5 μg of total RNA was used. RT-PCR was performed with the Thermoscript™ RT-PCR System (Invitrogen) as per the manufacturer's instructions. For the PCR portion of RT-PCR, a hotstart was employed at which time *Taq* DNA polymerase was added and the cycling reaction was started. Cycling temperatures and lengths varied with the primer sets.

2.2.3 5'/3' RACE

5' and 3' RACE was performed with the Marathon™ cDNA amplification kit (Clontech) as per the manufacturer's instructions (see **Appendix 1, Figure A6** for details). The

source RNA was 1 μg of human placental mRNA (the control RNA supplied with the kit). The cDNA synthesis primers were either oligo(dT) or *CECR2* R7 and/or R20 for 5' RACE and *CECR2* F1 for 3' RACE (**Figure 6**). In addition, 5' and 3' RACE was performed on Marathon Ready cDNA (Clontech) sourced from fetal brain and heart as per the manufacturer's instructions. Products were gel purified, subcloned into pGEM-T or pGEMT-Teasy (Promega), and sequenced on ABI 373 or 377 automated sequencers.

2.2.4 Genomic "RACE"

To PCR amplify a region of BAC 357F7 contained within a sequencing gap, a variation of the standard RACE protocol was used (see **Appendix 1**, **Figure A7** for details). Instead of double stranded cDNA as the source DNA, *Pvu II* digested genomic DNA was used. Other than the source DNA, the rest of the procedure followed the Marathon™ cDNA amplification kit (Clontech). Primers used in the primary and nested PCR reactions were *CECR2 R21* and *CECR2 R22* respectively (**Figure 9**). The resulting product was subcloned into pGEM-Teasy (Promega) and sequenced on an ABI 377 automated sequencer. This product was used to screen a mini-library constructed from digested BAC 357F7 to clone the region surrounding exon 1 of *CECR2*

2.2.5 Single fly PCR

Genotyping of individual flies was performed based on the method described by Saiki *et al.* 1998. Individual flies were homogenized with a pipette tip in a 0.6 mL eppendorf tube. 50 μL of "squishing buffer" (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 25 mM NaCl, 250 μg/mL proteinase K) was then expelled into the tube and the flies were

incubated at 56 °C for 30 to 60 minutes. One µL of the mix was then used in a PCR reaction.

The P element location and orientation was confirmed by PCR with different vector and *CG10115* intron 1 primer combinations. P element specific primers were SK08 and 11rpt5 (**Figure 28**). These primers were used in combination with Dm CECR2 F4 and R4.

To confirm if a deletion was present, a multiplex PCR reaction was run consisting of a positive control primer set (Su(var) 3-7 Fwd/Rev) and a primer set spanning the P element insertion site (Dm CECR2 F4/R4). Cycling conditions were as follows: hotstart (94 °C, 90", 80 °C hold followed by the addition of *Taq*) and then 61 °C, 20", 72 °C 45", 94 °C 10" for 35 cycles.

2.2.6 CECR2/Cecr2 cloning

CECR2 was cloned by RT-PCR and RACE as described in section **2.11**. Products were typically subcloned into pGEM-Teasy (**Appendix 1, Figure A1**) for sequencing and manipulation. Larger constructs were built by either cutting segments with enzymes followed by ligation to other segments, or by reamplification of the cloned inserts by PCR (see *2.2.7*).

2.2.7 Expression constructs

The expression constructs built for *CECR2* were generated by PCR of existing cDNA clone fragments in pGEM-Teasy using primers at the 5' and 3' ends of the coding region that contained engineered restriction enzyme recognition sites within them. The

cDNA was split in half at the site of a native *Bam*HI site over which primers were designed such that they incorporated the *Bam*HI site (**Appendix 1, Figure A3**). *CECR2* was amplified in two halves overlapping the *Bam*HI site. This allowed PCR products to be cut with the restriction enzymes and then ligated together into the modified pMIB/V5-HisA (see **Appendix 1, Figures A4 and A5**) expression vector.

2.3 Mouse genetic engineering

2.3.1 BAC transgenic mouse creation

Transgenic mice were created to carry extra copies of human BAC 357F7 (AC004019). BAC DNA was prepared as described in section *2.1.3*. BAC 357F7 DNA (~ 2.5 ng/μL) was injected into the male pronucleus of fertilized mouse oocytes (strain FVB/n) by Dr. Peter Dickie (HSLAS) as per (Hogan *et al.* 1994). Injected oocytes were then implanted into pseudopregnant FVB/n females. Pseudopregnant females were generated via mating to vasectomized males (performed by Dr. Peter Dickie).

Founder animals were initially screened for presence of the BAC transgene by Southern blot on DNA extracted from a tail clipping of weaned pups using a *CECR2* probe. Later, a PCR based assay was developed (see *2.6.1*).

2.3.2 Genetrap knockout mouse creation (chimera)

Mouse ES cell line CT45 (strain: 129P2/Ola HSD) containing the pGT1 splicetrap vector (Wilson *et al.* 1995) within *Cecr2* was obtained from Dr. Wendy Bickmore (MRC Human Genetics Unit, Edinburgh). ES cell line CT45 was grown by Dr. Roseline Godbout (Cross Cancer Institute, University of Alberta) under the following conditions:

cells were grown on plates pretreated with 0.1% gelatin in ES cell culture medium (DMEM with 2 mM L-glutamine, 100 μM β-mercaptoethanol, 0.1 M MEM non essential amino acids solution (Invitrogen), 200 μg/mL G418, 1000 U/mL LIF, 15% fetal calf serum (Wysent), and appropriate concentrations of penicillin-streptomycin). Cells were fed and split on alternating days. Dr. Peter Dickie then performed blastocyst injection of CT45 ES cells into blastocysts from strain C57BL/6J (Hogan *et al.* 1994). Injected blastocysts were implanted into pseudopregnant C57BL/6J females. As C57BL/6J animals have a black coat colour and 129P2/Ola animals have a yellow-agouti coat color, chimeric animals will be multi coloured while non-chimeric animals will be black.

2.4 Mouse colony maintenance

2.4.1 Housing conditions

All mice in this study were housed in the animal facility at HSLAS (Health Sciences Lab Animal Services) at the University of Alberta. Animals were housed in standard filter-top cages with free access to water and food (Purina 5001). Animals were kept under a standard breeding light cycle of 12 hours light followed by 12 hours dark at a room temperature of 21°C.

2.4.2 Mating

Mating of animals was determined by vaginal plug testing. From 1 to 3 females were added to the cage of a male animal in the afternoon. The following morning females were checked for the presence of vaginal plugs. The presence of a plug was treated as 0.5 days post-coitum (dpc). Plugged females were subsequently removed to individual cages.

2.4.3 Euthanasia

Mice were euthanised in a standard CO_2 chamber. On rare occasions, cervical dislocation was performed. Death was confirmed by a pain response test, which consisted of pinching of a paw.

2.5 Mouse embryo manipulation

2.5.1 Xgal Staining

The pGT1 genetrap vector is a splicetrap vector containing a β-galactosidase reporter gene and a neomycin resistance selectable marker, which splices in-frame to the mRNA of the locus in which it is inserted (in this case *Cecr2*). ES cells containing the genetrap in the correct reading frame of the gene into which it is inserted will display resistance to G418 and express β-galactosidase under the control of the locus into which it is inserted. Mouse embryos containing the pGT1 were stained with Xgal to determine the expression pattern of the *Cecr2* transcript. The procedure was as follows: embryos were removed from timed pregnant females (CO_2 euthanized) at time points between 9.5 and 15.5 dpc, placed in phosphate buffered saline (PBS), and the extraembryonic membranes dissected with the use of a stereomicroscope. Extra-embryonic membranes were saved for DNA extraction and genotyping. Embryos were fixed for at least one hour in 4% paraformaldehye/PBS at 4 °C. Fixed embryos were then rinsed in 1X PBS three times for 5 minutes each, followed by three washes of 30 minutes each in LacZ wash solution (1X PBS pH 7.2, 2mM $MgCL_2$, 0.01% deoxycholic acid, 0.02% IGEPAL). The embryos were then stained in the Xgal stain solution (1X PBS pH 7.2, 2 mM $MgCL_2$,

0.01% deoxycholic acid, 0.02% IGEPAL, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 1 mg/mL Xgal) at 37 °C until the desired staining intensity was reached (several hours to overnight). After staining, embryos were washed twice for five minutes each in 1X PBS to remove the stain solution. Embryos were then dehydrated through a methanol gradient (25, 50, 75, 100%) and stored in 100% methanol until photography or paraffin embedding.

2.5.2 Cartilage Staining

Mouse embryos were stained with the cartilage specific dye Alcian Blue 8GX (Sigma) as follows: pregnant mice were euthanised by CO_2 at 14.5 dpc and the embryos dissected out in phosphate buffered saline with the use of a dissecting scope. Extra-embryonic membranes were removed from each embryo and saved for DNA extraction and genotyping. All following manipulations were carried out in 15 mL polypropylene tubes on a rotary shaker, at room temperature. Embryos were fixed for a minimum of 2 hours in Bouin's fixative (53% ethanol, 10% formaldehyde, 6.7% glacial acetic acid, 0.45% w/v picric acid). Following fixation, embryos were washed (70% ethanol, 0.01% NH_4OH) six times for one hour each, with at least one wash going overnight or until the embryos turned pale white. The following day the embryos were rinsed twice, for one hour in rinse solution (1 part glacial acetic acid, 19 parts 70% ethanol). Embryos were then stained overnight in the staining solution (0.05% Alcian blue 8GX [Sigma], 1 part glacial acetic acid, 19 parts 70% ethanol). The following day the embryos were washed twice for one hour each in the rinse solution, followed by two one-hour washes in 100% methanol. Finally, the embryos were cleared in the clearing solution (2 parts benzyl

benzoate, 1 part benzyl alcohol). The embryos were photographed while submerged in the clearing solution. Long term storage of embryos was in the clearing solution.

2.6 Mouse genotyping

Weaned mice and mouse embryos were genotyped by PCR of tail DNA and extra-embryonic membrane DNA respectively.

2.6.1 Human BAC transgene detection (CECR2 containing)

The human BAC 357F7 transgene was detected by a multiplex PCR reaction containing a human specific *CECR2* primer set (*CECR2* F26 UTR/R28 UTR), and a mouse control primer set (pGT1 F1/R1, from the *EN2* gene). The reaction contained 10 mM Tris-HCl pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, 0.02 mg/mL BSA, 250 μM dNTPs, 2.5 U Taq DNA polymerase and 20 pmol of each primer. One μL of previously isolated tail or embryonic sac DNA was used in each reaction. Reactions were cycled as follows: a "hotstart" of 94 °C for 1'30", followed by holding at 80 °C at which time the *Taq* was added. After the addition of *Taq*, a two-step PCR was cycled: 68 °C for 45", 94 °C for 15" for a total of 35 cycles. Reactions were then separated on a 2% agarose gel.

2.6.2 Mouse pGT1 genetrap detection

Mice and embryos in the *Cecr2* knockout experiment were genotyped by a multiplex PCR reaction containing 5 primers. The 5 primers were: *Sry* Fwd/Rev, *Cecr2* 7i.Fwd/Rev and pGT1-R4. *Sry* Fwd/Rev will amplify a portion of the chromosome Y specific gene *Sry*, which identifies male animals. *Cecr2* 7i.Fwd/Rev will amplify the

wildtype *Cecr2* intron 7 allele, while *Cecr2* 7i.Fwd/pGT1-R4 will amplify the mutant *Cecr2* intron allele which contains the pGT1 genetrap vector. The reaction contained 10 mM Tris-HCl pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, 0.02 mg/mL BSA, 250 μM dNTPs, 2.5 U *Taq* DNA polymerase and 20 pmol of each primer. 1 μL of previously isolated tail or embryonic sac DNA was used in each reaction. Reactions were cycled as follows: a "hotstart" of 94 °C for 1'30", followed by holding at 80 °C at which time the *Taq* was added. After the addition of *Taq*, the reactions were cycled as follows: 62 °C for 20", 72 °C for 45", 94 °C for 15" for a total of 35 cycles. Products were separated on a 2% agarose gel.

2.7 *Cecr2* knockout mice rescue experiments

2.7.1 Human BAC transgene rescue

The human *CECR2* transgene (BAC 357F7) was crossed into the pGT1 genetrap line to determine if it would rescue the mutant phenotype. The two *CECR2* transgene lines (Lines 1 and 10) from the inbred FVB/n strain were crossed into $Cecr2^{+/pGT1}$ mice from a 129P2/Ola:BALB/c mixed strain to generate F1 animals that were used as stud mice in the rescue experiment (**Figure 27**). Genotyping was performed on tail DNA (as described in sections *2.6.1* and *2.6.2*). Two individual PCR reactions were performed for each DNA sample (one for *Cecr2* genotype and the other for *CECR2* presence/absence). Typically, male individuals from each rescue line (1 and 10) that had the genotype $Cecr2^{+/pGT1}$, +*CECR2* TgN were used to mate to $Cecr2^{+/pGT1}$ female animals. Embryos generated from these matings were collected between 12 and 14 dpc, and were genotyped and analyzed for exencephaly. The complimentary matings (with female F1 $Cecr2^{+/pGT1}$,

+*CECR2* TgN animals mated to *Cecr2*$^{+/pGT1}$ males) were also performed with any F1 females that contained the appropriate genotype. Penetrance of exencephaly was compared between *Cecr2*$^{pGT1/pGT1}$ and *Cecr2*$^{pGT1/pGT1}$, +*CECR2* TgN animals.

As the F2 animals in this experiment had a genetic background originating from three strains (BALB/c, FVB/n, and 129P2/Ola), a control experiment was also performed in order to determine exencephaly penetrance on this new background. This control experiment was carried out the same as the experimental one with one exception: the F1 stud males (littermates of the experimental stud males) had the genotype *Cecr2*$^{+/pGT1}$, while maintaining the same mixed genetic background (129P2/Ola:BALB/c:FVB/n) as the experimental stud males. Some genotyping later in the project was performed by Tanya Ames.

2.7.2 Folic acid rescue

This experiment was carried out on female mice of genotype *Cecr2*$^{+/pGT1}$ mated to males of genotype *Cecr2*$^{+/pGT1}$. The strain background was 129P2/Ola:BALB/c mixed. Upon detection of a plug, females were put on a daily regimen of intraperitoneal injection of PBS (control group) or Folic acid/PBS (experimental group). The experimental group of females were injected with 0.5 mg/mL folic acid (dissolved in PBS and 0.22 μM filtered) to a final concentration of 10 mg/kg. Females were weighed before plug testing to determine the appropriate injection volume, and every few days thereafter to accommodate natural weight gain. At 12-14 dpc when the females were visibly pregnant, embryos were harvested, genotyped and analyzed for exencephaly.

2.8 RNA manipulation

2.8.1 Total RNA isolation

RNA was isolated from various human and mouse tissues with the use of Trizol (Invitrogen). Frozen tissues were homogenized in the presence of Trizol (1 mL per 100 mg of tissue) with an electronic tissue homogenizer (Janke & Kunkel). Samples were allowed to sit for 5 minutes at RT, at which time chloroform was added (0.2 mL per 1 mL of Trizol) and the samples were vortexed. Samples were allowed to sit at RT for 3 minutes followed by centrifuging at 14,000 rpm for 15 minutes at 4 °C. The aqueous portions were removed to new tubes and to them 1/10 volume of 3M NaOAc and 1 volume of cold 2-propanol was added. Samples were left at RT for 10 minutes and then centrifuged at 14,000 rpm for 12 minutes at 4 °C. The subsequent pellet was washed in 75% ethanol (made with DEPC water) and recentrifuged for 5 minutes at 14,000 rpm at 4 °C. The pellet was dried and resuspended in DEPC H_2O.

2.8.2 RNA electrophoresis

RNA samples were electrophoresed in a formaldehyde agarose gel composed of 1.2% agarose, 1.85% formaldehyde and 1X MOPS (0.42% MOPS, 5mM NaOAc, 1mM EDTA, pH 7). Each RNA sample was placed in a sample buffer of 6.6% formaldehyde, 1X MOPS and 50% formamide. Samples were heated at 60 °C for 10 minutes to remove any secondary structure from the RNA prior to loading, followed by the addition of 1/10 volume of loading dye (1 mM EDTA pH 8, 0.25% bromphenol blue, 0.25% xylene cyanol and 50% glycerol). Samples were loaded and electrohoresed at 100 volts until the dye had moved 2/3 of the way through the gel. The gel was then stained in DEPC H_2O

and ethidium bromide (2.5 μg/mL) for one hour, destained in DEPC H_2O for one hour and photographed.

2.8.3 Poly-A^+ (mRNA) RNA isolation

PolyA$^+$ RNA was purified from total RNA previously isolated from mouse 12.5 - 15.5 dpc embryos (see section *2.8.1*). The PolyATract® mRNA Isolation System IV (Promega) was used to isolate the mRNA. The entire total RNA aliquot previously isolated from a single 14.5 dpc mouse embryo was used as the starting input for mRNA purification. The procedure was as per the manufacturer's instructions. The mRNA was eluted into DEPC H_20 for further analysis.

2.9 Probe labeling

DNA templates were radioactively labeled for use in Southern or Northern blot hybridization. Probes were generally double stranded DNA isolated from agarose gels. DNA probes were labeled either by Klenow or PCR based kits from Ambion (below).

2.9.1 Klenow labeling

Up to 25 ng of double stranded DNA was labeled with the Strip-EZ™ DNA probe labeling kit (Ambion) as per the manufacturer's instructions. Briefly, up to 25ng of dsDNA was denatured and added to random decamers (or hexamers), buffer, modified dCTP (Ambion), α-^{32}P dATP, and Klenow enzyme as per the kit instructions. After labeling, the probe was boiled for 2 minutes and added to the top of a Sephadex G50 (Pharmacia) column generated inside a 1 mL syringe, and centrifuged at 1,200 rpm for 2

minutes in a clinical centrifuge to separate the unincorporated nucleotide from the labeled probe.

2.9.2 PCR labeling

Alternatively, double stranded DNA was PCR labeled with the Strip-EZ™ PCR labeling kit (Ambion) as per the manufacturer's instructions. Briefly, up to 50 ng of linearized template was added to PCR buffer, dNTPs, α-^{32}P dATP, sense and antisense primers as per the kit instructions. *Taq* DNA polymerase was added after a hotstart and the reactions were cycled for 35 cycles of: 95°C 30", 55°C 20", 72°C 60" (for products 1kbp or less). After labeling, the probe was boiled for 2 minutes and added to the top of a Sephadex G50 (Pharmacia) column and centrifuged at 1,200 rpm for 2 minutes to separate the unincorporated nucleotide from the labeled probe.

2.10 Hybridizations

2.10.1 Southern

All Southern blots (Southern 1974) were performed on DNA bound to Genescreen^{+} (Dupont) charged nylon membrane. DNA was transferred to membrane via a capillary stack with alkaline buffer, as per the manufacturer's instructions. Probes were labeled as described in section *2.9.1* and *2.9.2*. Probes were hybridized to blots overnight at 65 °C in rotating bottles in a hybridization oven (Tyler) in hybridization buffer (10% SDS, 1 M EDTA, 1 M $NaPO_4$, 5X Denhardt's{0.1% FicoII 400, 0.1% polyvinylpyrrolidone, 0.1% BSA fraction V}). Blots were washed once at room temperature in low stringency wash (2X SSC {300 mM NaCl, 30 mM sodium citrate},

0.2% SDS) followed by high stringency washes at 65°C (0.2X SSC, 0.1% SDS) until background was deemed low enough by Geiger counter. Blots were then sealed in plastic bags and exposed to Xray film.

2.10.2 Northern

Northern blots were purchased from Clontech, and contained mRNA from multiple human tissues. Probes were labeled as described in section **2.9.1** and **2.9.2.** Probes were hybridized to blots overnight at 42 °C in rotating bottles in a hybridization oven (Tyler) in hybridization buffer (50% formamide, 5X SSPE {0.25% SDS, 2.5 mM EDTA}, 2% w/v SDS, 10X Denhardt's, and 100 µg/mL sonicated salmon sperm DNA). Following hybridization, blots were washed once at room temperature in low stringency wash (2X SSC, 0.1% SDS) and then at 42-50 °C in high stringency wash (0.1X SSC, 0.1% SDS) until background was deemed low enough by Geiger counter. Blots were then sealed in plastic bags and exposed to Xray film.

2.11 Protein manipulation

2.11.1 InsectSelect™ protein expression

Insect cell lines SF9 (*Spodoptera frugiperida*, pupal ovary), S2 (*Drosophila melanogaster*, late stage embryo), and KC1 (*Drosophila melanogaster*, embryo) were used to express recombinant GCN5 and CECR2 in conjunction with the InsectSelect™ protein expression system from Invitrogen. The pMIB/V5 vector adds a honeybee mellitin secretion signal to the amino terminus of the protein such that the expressed

protein should be secreted into the media. In addition the vector adds a V5 epitope and a 6X Histidine tag on the carboxy terminus for identification and purification purposes.

2.11.2 Transfection of insect cell lines with InsectSelect™ expression vectors

Insect cell culturing and transfections were performed by Heather McDermid. Transfections were carried out in standard 6 well tissue culture plates. S2 cells were grown in CCM3 media (Hyclone) while SF9 and KC1 cells were grown in ESF921 media (Expression Systems). Cells were grown in T25 flasks at 27°C and allowed to grow until the target density (SF9: 2-3 X 10^6 cells/mL, S2 and KC1: 4-5 X 10^6 cells/mL) was achieved. At this time cells were moved into 6 well plates. For SF9, 2-3 X 10^6 cells were added to 1 mL of Graces's medium and allowed to attach. For S2 and KC1 the same procedure was carried out, with the exception that 4-5 X 10^6 cells were transferred.

As cells were attaching, ~1 μg of expression construct DNA (resuspended in sterile water) was added to 1 mL of Grace's insect medium (Sigma) in a 1.5 mL eppendorf tube and mixed. Ten μL of Cellfectin (Invitrogen) was then added to the DNA/media mixture and mixed. This mixture was allowed to stand for 30 minutes to allow DNA/liposome complexes to form. After 30 minutes, the media was removed from the cells and it was replaced with the DNA/media/Cellfectin mixture. After 3-4 hours, 1.5 mL of fresh media (CCM3 or ESF921) was added to each well. The plates were sealed with Parafilm™ and the cells were grown at 27°C for 48-72 hours and either harvested to look for transient protein expression, or were transferred to T25 flasks in preparation for selection with Blasticidin, which selects for integration of the vectors.

2.11.3 Generation of stable recombinant insect cell lines

Blasticidin was used to select for vector integration and generation of stable cell lines expressing recombinant CECR2 or GCN5. Cells were split from the 6 well tissue culture plates into T25 flasks and allowed to grow overnight. The following day the medium (CCM3 or ESF921) was removed along with any unattached cells and replaced with fresh media containing Blasticidin (70 μg/mL). Cells were grown for 5 days and then passaged, and fresh media with Blasticidin was added. During the selection process, untransfected Sf9 and S2 cell lines were subjected to Blasticidin treatment to confirm that it was lethal to these cells. Cells were subjected to selection levels (70 μg/mL) of Blasticidin for 2 passages and then moved to maintenance selection in Blasticidin (10 μg/mL) for propagation.

2.11.4 Recombinant protein isolation and purification

Cells were harvested by resuspending attached cells into the media with a pipette, followed by centrifugation in 15 mL polypropylene tubes in a Sorvall SH3000 swinging bucket rotor at 3000Xg and 4°C. The media was removed and saved for processing, and the pellets frozen at –70°C until protein isolation was to continue. Protein isolation consisted of resuspending pellets in 250 μL of lysis buffer (150 mM NaCl, 1% IGEPAL, 50 mM Tris pH 8.0), which was then passed through a 26 gauge needle three times. The mixture was centrifuged at 14,000 rpm in a microcentrifuge to pellet cell debris. The supernatant was removed to a new tube and was either frozen at –70°C until further processing or was taken directly into purification via nickel-agarose affinity chromatography. The original cell media was concentrated approximately 10-15 fold in

Amicon® Ultra-15 centrifugal filter tubes. These concentrators had a molecular weight cut-off of 10 kDa and were centrifuged at 4°C in a Sorvall SH3000 swinging bucket rotor at the recommended velocity until the media was concentrated from 3 mL to approximately 200 µL. This generally took 20-30 minutes. This concentrated protein solution was then either frozen at –70°C or a sample was immediately electrophoresed on a SDS-PAGE gel for analysis.

2.11.5 Protein electrophoresis

Proteins were separated on standard Laemmli Tris-glycine SDS-PAGE gels, with the exception of HAT assay gels, containing stacking and resolving portions. The resolving gel (acrylamide 7.5 -15%, 0.375 M Tris-HCl pH 8.8, 0.1% SDS) was polymerized with the addition of ammonium persulfate and TEMED to final concentrations of 0.05% and 0.03% respectively. The resolving gel was overlaid with water or n-butanol to enhance polymerization. After polymerization, the top of the resolving gel was washed and any residual water or butanol was removed. A stacking gel (4% acrylamide, 0.125 M Tris-HCl pH 6.8, and 0.1% SDS) was poured on top of the resolving gel. The gel was polymerized as per the resolving gel.

Protein samples were added to an equal volume of 2X "cracking" buffer (0.125 M Tris-HCl pH 6.8, 5% SDS, 10% sucrose, 10% β-mercaptoethanol), boiled for 2 minutes and loaded into the gel. Gels were electrophoresed at a constant current of 25 mA per gel in Tris-glycine running buffer (0.025 M Tris, 0.192 M glycine and 0.1% SDS) until the desired separation was achieved. The gels were subsequently removed and either stained with Coomassie or silver, or electroblotted to nitrocellulose membrane.

2.11.6 Polyacrylamide gel staining

SDS-PAGE gels were stained either with silver, or Coomassie Blue. Some gels were stained with Coomassie as per (Sambrook *et al.* 2001), while others were stained with the Silver Stain Plus™ kit (Biorad) as per the manufacturer's instructions.

2.11.7 Protein transfer to membrane by electroblotting

SDS-PAGE gels were electroblotted to nitrocellulose in a Hoefer transfer apparatus overnight at a constant current of 100 mA in Towbin transfer buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, 10% methanol). Membranes were then directly taken into the western blotting protocol.

2.11.8 Western blotting

Western blots were performed as follows: membranes were blocked for one hour at room temperature in blocking solution (5% non-fat milk dissolved in 1X Tris buffered saline with 0.1% Tween-20 {TBST}) on a rocking platform. This was followed by three 5 minute washes in 1X TBST. The primary antibody was diluted (typically 1:3000 – 1:5000) in 2.5% milk/TBST. Incubation of the primary antibody was for 1 hour at room temperature on a rocking platform. This was followed by three 5 minute washes as before. The secondary antibody was performed as per the primary antibody. The secondary antibody was anti-mouse conjugated to horseradish peroxidase, and was typically diluted 1:5000. Following the secondary antibody incubation, the blot was washed for three 5 minute washes in TBST. The blot was developed with the Amersham

ECL detection kit as per the manufacturer's instructions. The blot was then exposed to X-ray film.

2.11.9 Histone acetyltransferase assay

The HAT assay consisted of a reaction containing radiolabeled acetyl-CoA, histones, buffer and input protein, which were incubated and then run on a high-Tris/Urea SDS PAGE gel. The reactions consisted of 25 μg recombinant protein, 15 μg calf thymus histones (Sigma), 0.5 μCi ^{3}H-Acetyl-CoA in 1X reaction buffer (1 M DTT, 0.1 mM EDTA, 50 mM Tris pH 8.0, 10% glycerol, 0.5% IGEPAL, plus 1 complete protease inhibitor tablet (Roche) per 5 mL of buffer). The reaction was started with the addition of histones and then incubated for 30 minutes at either 30 °C or 37 °C. After incubation, the reactions were stopped by the addition of 2X SDS PAGE cracking buffer, followed by heating at 95 °C for 2 minutes. These reactions were then run on a high Tris/Urea SDS PAGE gel (separating gel = 26.7% w/v acrylamide, 0.35% w/v bis-acrylamide, 1 M Tris pH 8.0, 0.55 mM NaCl, 0.14% SDS v/v, 50% urea w/v; stacking gel = 6.7% acrylamide, 0.09% bis-acrylamide, 0.17 M Tris pH 6.8, 50% urea w/v, 0.14% SDS [acrylamide stock = 60% acrylamide:0.8% bis-acrylamide]). The gel was then fixed (in 30% methanol:10% glacial acetic acid) for 30 minutes, washed in water twice for 30 minutes and subsequently stained (as per *2.11.5*). The fixed and stained gel was then subjected to fluorography: the stained gel was soaked in En^{3}Hance fluor (NEN) for 1 hour at room temperature on a rotary shaker. The fluor was then removed and the gel soaked in cold H_2O for 1 hour to precipitate the fluor. The gel was then dried and exposed to Xray film for 5-14 days.

Done in duplicate, the second set of reactions were subjected to scintillation counting as opposed to electrophoresis and fluorography. After the reactions were incubated they were stopped by spotting on Whatman P81 filters (Millipore) and allowed to dry. These filters were subsequently washed in 0.2 M $NaCO_3$ pH 9.2 for 20 minutes, three times. The filters were then dried and subjected to scintillation counting for ^3H in scintillation fluor.

3. RESULTS

3.1 Gene mapping in the CESCR

Computer aided sequence analysis was undertaken to identify genes within the cat eye syndrome critical region (CESCR). A contig of BAC clones had already been created (Johnson *et al.* 1999) and was in the process of being sequenced at the University of Oklahoma, as part of the Human Genome Project. My search began with unfinished sequence of the BAC clone 77H2 (*AC000052*) and then later an overlapping BAC clone 357F7 (*AC004019*) (**Figure 2**). These sequences were used for BLAST searching (Altschul *et al.* 1990) and gene prediction with GENSCAN (Burge and Karlin 1997). Prior to gene identification all sequences had repetitive elements removed with Repeatmasker (see **Appendix** for internet address). The repeatmasked sequence was queried against various Genbank databases (EST {expressed sequence tag}, HTGS {high throughput genomic sequence} and NR {non-redundant}) to search for conserved or expressed sequences.

3.2 Cloning and characterization of SLC25A18, a mitochondrial glutamate transporter

A search of the dbEST database yielded a cluster of ESTs, which appeared to be the 3' UTR of an uncharacterized gene near the telomeric end of both BACs (**Figure 2, 3**). One of these ESTs (EST 28949; *R40846*) was obtained and sequenced. EST 28949 was from an infant brain cDNA library and contained a reported insert of 1845bp. When comparing the sequence of EST 28949 to BAC 77H2, it was clear that 10 exons were contained within the cDNA. This allowed orientation of the unfinished genomic sequence

contigs and allowed further EST searching and gene prediction with GENSCAN. GENSCAN predicted a gene comprised of 11 exons, which included the 10 exons identified from the EST 28949 sequence, in addition to one upstream exon. EST searching confirmed all 11 exons, although it is of note that several ESTs appear to have spliced out exon 2. Analysis of the putative transcript identified an ORF of 945 bp, and a predicted protein of 315 aa. The putative start codon is in exon 3, such that alternative splicing of exon 2 should not affect the encoded protein.

Hybridization of EST 28949 to commercial Northern blots showed that *SLC25A18* is expressed in brain and liver (**Figure 4**) out of the eight tissues represented on the two blots. Three alternative transcripts are evident, one in brain (~2.2 kb) and two in liver (~2.0 and 3.6 kb). All three transcripts appear to be found in both adult and fetus, albeit at a lower level in fetus (3.6 kb transcript is clearly visible on the fetal blot autoradiogram). The two liver transcripts appear to differ in size by at least 1.5 kb. It seems most plausible that this part of the mRNA would be comprised of UTR, as this family of mitochondrial carriers are usually around 300 aa, yet ESTs have not been observed with a longer, or novel 5' or 3' UTR. At this time, the composition of either of the liver transcripts is not clear. Exon 2 of *SLC15A18* is only 63 bp, the removal of which would not likely be resolvable as a second, smaller band on a Northern blot. Thus, the single signal in brain is likely composed of the two transcript variants, with and without exon 2.

Analysis of the predicted protein sequence of *SLC25A18* yielded significant information about the protein. A hydropathy plot suggested 6 transmembrane domains (not shown). A BLAST search of the non-redundant protein database returned many

proteins from the mitochondrial carrier superfamily, such as Citrin (*SLC25A13*) and Aralar (*SLC25A12*). These are proteins found in the outer and inner mitochondrial membranes, and transport solutes in and out the various mitochondrial compartments (reviewed in Palmieri *et al.* 2000). Thus, at the time of cloning, analysis suggested that *SLC25A18* was likely a mitochondrial solute carrier, but it was unknown as to which solute it transported.

3.3 Localization and Northern analysis of ATP6E

The epsilon subunit of vacuolar ATPase (*ATP6E*) had been previously mapped to the CESCR (Baud *et al.* 1994), but its precise location was not known. Analysis of BACs 77H2 and 357F7 sequence showed that *ATP6E* was truncated on BAC 77H2 and absent from BAC 357F7. *ATP6E* is oriented telomere to centromere and its 3' end abuts closely (1.3 kbp) with the 3' end of *SLC25A18*, which is oriented centromere to telomere (**Figure 2, 3**). Baud *et al.* had reported RT-PCR of this transcript on various tumour cell lines, but not on any normal tissue. I used the 61EW cDNA probe (Baud *et al.* 1994) as a probe on a human multi tissue Northern blot. A single transcript of approximately 1.45 kb was observed in all 8 tissues contained on the blot (**Figure 5**). When compared to the actin control, expression of *ATP6E* appeared to be relatively uniform across the 8 tissues. Interestingly, Baud *et al.* found two distinct RT-PCR products in 2 of 6 different tumour cell lines tested. The possibility of alternative splicing cannot be ruled out, although the Northern analysis suggests that potential alternative transcripts either are not found at levels detectable by Northern blotting, or these putative events are not found in the 8 tissues tested.

3.4 Identification of CECR2, a putative chromatin remodeling protein

CECR2 was identified in the same manner as SLC25A18, through a combination of gene prediction and BLAST searches of dbEST using the BAC 77H2 and BAC 357F7 sequences. Several exons were predicted by GENSCAN, and the single pass sequence of one EST (852605, *AA663110*) matched four of the predicted exons. This EST was sequenced and confirmed the GENSCAN predicted exons that would later be defined as exons 2 through 13' (**Figure 6**). Confirmation of the exons allowed new primers to be designed and subsequently allowed RT-PCR, and 5'/3' RACE.

The initial 3' RACE yielded a product of approximately 2400 bp from fetal brain (**Figure 6**). This product was subsequently subcloned into pGEM-Teasy and sequenced. Comparing the cDNA sequence to BAC 77H2 sequence, exons 10 through 17 were confirmed (**Figure 6**). However, it was clear that the end of the 3' RACE product was not a bona-fide 3' end of the mRNA as there was no stop codon, polyadenylation signal or poly-A tail identified. However, identification of exons 11 through 17 allowed the unfinished 77H2 contigs to be oriented and a new gene prediction to be performed. This time GENSCAN predicted a partial gene with 18 exons, but lacking an initial exon. All predicted exons matched sequence from previously identified exons, via sequenced ESTs or RACE products, with the exception of two more 3' exons, which turned out to be exons 18 and 19. Exon 19 only has one codon before the stop codon and subsequent untranslated region. BLAST searches of dbEST then identified several ESTs, which mapped to exons 17, 18 and 19, the latter of which had legitimate polyadenylation signals. This, in combination with the fact that exon 19 appeared to be almost entirely

untranslated, suggested that it was the terminal exon of *CECR2*. In addition, BLAST of dbEST showed three separate clusters of 3' ESTs within exon 19 (**Figure 6**). The first two clusters contained canonical polyadenylation signals (AATAAA and ATTAAA) and corresponded to exon sizes of 1111 and 1307 bp. These two alternative polyadenylation events produced distinct 3' ends designated exon 19 and 19'. A third EST cluster was also confirmed to be part of the *CECR2* transcript producing a 5313 bp exon 19". This long version of exon 19" was confirmed to be part of the transcript by Northern blot (data not shown) and by using a cDNA synthesis primer within this region to perform RT-PCR on various tissues between exons 7 and 9.

Five prime RACE was performed on oligo-dT primed fetal brain cDNA with nested primers *CECR2* R7 and R11 (exon 10). Various products were obtained and these were subcloned. The clones were hybridized with EST 852605 (exons 2-13') to determine which ones were legitimate. One of the clones contained an approximately 1000bp insert, which upon sequencing confirmed exons that matched EST 852605 between exons 2 and 10, but had novel sequence at the 5' end (**Figure 6**). At the time of this experiment, the novel *CECR2* 5' exon sequence did not match any sequence on the completed BAC 77H2 or the partially sequenced overlapping clone BAC 357F7. A probe was generated from the most 5' 200 bp (*CECR2* exon 1) of the R11/AP2 5' RACE product and was hybridized to a Southern blot containing the genomic clones from the CESCR (clone locations shown in **Figure 2**). This probe hybridized to the overlapping BAC clones 872B4, 233A2 and 357F7, but not 77H2 (data not shown). This suggested that the novel 5' exon was not contained within BAC 77H2, but should be found in the 357F7 sequence when complete. Considering that exon 2 was found approximately

70kbp from the 5' end of BAC 77H2, this suggested that intron 1 was very large and likely why exon 1 was not contained within BAC 77H2. It was later determined that intron 1 is 106,779 bp in length. Due to the large size of *CECR2* intron 1, and the presence of regions of homology between human and mouse within this intron (Footz *et al.* 2001), it was queried against the EST database. We wanted to determine if other expressed sequence(s), possibly in the opposite direction of transcription from *CECR2*, were present within this intron. Other than one EST cluster which corresponds to a processed pseudogene of the TRF2-interacting telomeric RAP1 protein (Accn no. BC022428), no other EST clusters with more than 2 members were found in the telomere to centromere orientation. It seems unlikely that an independent transcription unit resides within *CECR2* intron 1.

Having shown that the most 5' 200 bp from the R11/AP2 5' RACE product mapped to BAC 357F7 but not 77H2 and was likely *CECR2* exon 1, I used its sequence to search the EST database. This novel (*CECR2* exon 1) sequence matched a 5' RACE sequence from a mouse ES cell genetrap clone, MMEST92 (*X98199*), suggesting it was indeed part of a transcript. At this point the corresponding mouse BAC clones had not yet been sequenced, so a comparison of the 5' RACE sequence with mouse genomic sequence was not yet possible. Also during this time, the sequence of BAC 357F7 was completed, yet did not contain the novel *CECR2* 5' RACE sequence. Since this novel 5' end did hybridize to a BAC 357F7 Southern blot, yet was missing from the final Genbank sequence of the clone, this suggested that the completed sequence of BAC 357F7 was missing a small region. From this information, I chose to clone the region of BAC 357F7 that contained the putative *CECR2* exon 1.

3.5 Identification of the *CECR2* initial exon and cloning/sequencing of a genomic sequencing gap

In order to clone genomic DNA from around the putative *CECR2* exon 1, a modified 5' RACE protocol was used on genomic DNA. A "RACE" library was constructed using *Pvu* II cut genomic DNA to yield relatively small, cloneable fragments. These fragments were then ligated to adapter linkers from a cDNA synthesis kit. Primers designed from the existing 5' RACE sequence were used for nested "RACE" PCR with linker specific primers and *CECR2* specific primers (*CECR2* R21 and R22, see **Figures 6, 9** for primer location). A product of approximately 90 bp was obtained (**Figure 7a**) and subcloned. Sequence of this product showed that it overlapped with both MMEST92 and the 5' end of the R11/AP2 5' RACE product, while having approximately 20 bp of unique sequence at the 5' end. This PCR product was radiolabeled and used as a probe on a Southern blot of BAC clones from the CES region. The probe hybridized to overlapping BAC clones 115F6, 357F7, 872B4 and 233A2, but not 77H2 (**Figure 7b;** see **Figure 2** for clone orientation). In addition, mouse BACs 555D9 and 369P18 hybridized to this probe. These mouse BACs were previously shown to map to the region of the mouse genome orthologous to human 22q11.2 (Footz 1999; Footz *et al.* 2001). This suggested that MMU EST92 (see section **3.4**) was likely from the mouse version of *Cecr2* and that the human 5' RACE products (R11/AP2 and R22/AP2) were legitimate. In addition, this also suggested that the isolate of BAC 357F7 that was sequenced at the University of Oklahoma contained a small deletion, or there was an error in closing the

sequence. As our isolate of BAC 357F7 hybridized to the 5' "genomic" RACE product (R22/AP2), I chose to identify the "deleted" region of BAC 357F7, clone and sequence it.

The first step of cloning the "deleted" region surrounding the *CECR2* first exon involved performing a Southern blot of 357F7, digested with various enzymes, using a *CECR2* exon 1 probe (R22/AP2). As shown in **Figure 8a**, *Sst*I cut within the probe region, and was thus not used for cloning. Other enzymes produced fragments that were too large for efficient cloning (>10 kbp). Therefore *Sst*I was chosen due to its reasonably sized fragment (~1450 bp) and the fact that it did not cut within the probe region. An *Sst*I cut, and 1-2 kbp size selected by electrophoresis, BAC 357F7 mini-library was then constructed in pBSII-SK$^+$ (**Figure A2** for vector info). Clones from this mini-library were hybridized with the R22/AP2 probe and several clones were isolated and sequenced. Sequence of these 1400 bp *Sst* I clones, anchored the fragment at its 3' end within BAC 357F7 at position 35,293 bp (of the Genbank file). It was determined that 779 bp of sequence from the 3' end of the *Sst* I fragment matched BAC 357F7. However, the remaining 5' sequence of the *Sst* I fragment still did not match any sequence on BAC 357F7 suggesting that it fell within the "deleted" region (**Figure 9**).

To clone the remainder of the deleted region, a second Southern blot of BAC 357F7 was performed with different restriction enzymes in order to identify a second fragment to subclone (**Figure 8b**). From this Southern blot it was determined that *Bam*HI was the best candidate as it produced a single fragment of a clonable size (~3.3 kbp). A mini-library approach was again taken in which BAC 357F7 was digested with *Bam*HI and fragments were shotgun subcloned into pBSK II$^+$ digested with *Bam*HI. Two of these clones were sequenced and the sequence obtained anchored the fragment on the 5' end of

the "deletion". However, it became apparent from the obtained sequence that this deleted region contained a large CpG island. The high GC content made the remaining region difficult to accurately sequence. New primers were designed on the 5' side of the deletion (357F7 F1 and F2) to allow PCR to determine the size of the missing region (**Figure 9**). The *Bam*HI subclones were used as a template for PCR and the primer pair 357F7 F2/R21 produced a product of approximately 575 bp, which suggested that the remaining deletion was small. This PCR product was sequenced multiple times with various primers and concentrations of DMSO in order to get accurate sequence in this region of high GC content. The entire deletion was determined to be approximately 960 bp, the distal 420 bp of which was intron, and the remainder proximal was likely *CECR2* exon 1 (**Figure 9**).

A BLAST of dbEST was performed with the sequence surrounding the putative *CECR2* exon 1, and the best matches included MMEST92 (a 5' RACE product from a mouse ES cell genetrap clone) in addition to three *Drosophila melanogaster* ESTs. This was further evidence that this sequence is expressed. While we cannot be sure of the transcription start site without further experiments such as primer extension, we can determine the translation start site with some certainty. Comparison with the mouse EST92 and the ORFs of the *Drosophila* ESTs identified a conserved ATG codon 126 bp proximal to the splice donor site. In addition, a stop codon is found 203 bp upstream with no other ATG codons in the intervening region (**Figure 9**). PCR primers (*CECR2* F23, 24) were designed upstream of this stop and RT-PCR was performed on *CECR2* R19 (exon 4) primed cDNA generated from DNase treated human kidney RNA. No products were obtained with any combination of the following forward and reverse primers: F23 or F24 and R20 (exon 4) or R22 (exon 1). Products were obtained with F22/R22 and

F22/R19 (data not shown). However the absence of product may be due to the high GC content of the sequence and subsequent failure of the PCR reaction. Further experimentation will be necessary to determine the true size of the 5' UTR and possible transcription start sites.

3.6 Northern analysis of *CECR2*

To determine the expression profile of *CECR2*, multi tissue Northern blots were hybridized with various *CECR2* probes. With a probe consisting of *CECR2* exons 1-10, there appear to be 3 predominant transcripts (10.0, 9.5 and 6.0 kb) in most of the tissues tested (**Figure 10**). There does appear to be some variation in the relative expression of the different transcripts in different tissues. Not all tissues appear to have *CECR2* expression: spleen, thymus, prostate, intestine and blood leukocytes did not appear to have appreciable *CECR2* expression based on the Northern blots. Also of note is that ovary and testis only contain a subset of the transcripts, with ovary expressing only the 10 kb transcript, and testis only expressing the 9.5 and 6 kb transcripts. A probe from exon 19" confirmed that it was part of the *CECR2* message, as the 10 and 9.5 kb bands hybridized (data not shown). This would suggest a transcript comprised of approximately 5 kb of ORF and 5 kb of untranslated region. It is unclear what role, if any, this large UTR may have (see **3.9**).

Due to the smearing on the blots, it is difficult to observe any smaller transcripts. Based on fetal retina EST 852605, which contains exons 1-13' (*CECR2b*), there should be a small transcript at approximately 1.4 kb but this is not seen on the blots. It is possible that this smaller transcript is expressed at levels undetectable by Northern blotting, or is

specific to a particular tissue not found on our Northern blots. Unfortunately due to the small amount of unique sequence in exon 13' that is not part of the larger transcripts, a probe specific to the small transcript could not be generated. Due to this problem, RT-PCR was used as an alternative method to determine expression of the small *CECR2b* transcript. It should also be noted that RNA *in situ* hybridization of *Cecr2* in mouse embryos was attempted, but no interpretable results were obtained. Various probes throughout the *Cecr2* coding region and untranslated region were used in RNA *in situ* hybridization experiments in conjunction with a *Pax1* positive control probe. The control probe always gave the appropriate signal, whereas the *Cecr2* probes never produced reproducible staining patterns. RNA *in situ* of *Cecr2* was not attempted further.

3.7 Identification of *CECR2* alternative transcripts

Based on the BLAST searches of the BAC357F7 genomic sequence, it was already clear that there were alternative polyadenylation events within *CECR2* exon 19, as 3 EST clusters were found, in addition to alternative splicing involving exon 13 (**Figure 6**). Due to these observations and the fact that the small *CECR2b* transcript (exons 1-13') could not be effectively resolved on a Northern blot, RT-PCR was used to determine the expression profile of the small *CECR2b* transcript. This led to the determination that more alternative splicing was present within the *CECR2* transcript than was originally thought, and a more detailed search for alternative splicing was undertaken. Primer pairs for RT-PCR were chosen to span at least one intron. All exons were tested where size permitted. In addition, cDNA synthesis primers were used to generate cDNA specific to exon 13'and 19" to determine splicing and expression of

CECR2 transcripts containing those 3' UTRs. These are the only two regions of the cDNA that are not shared by multiple putative transcripts, and as such the first strand cDNA should be specific to those particular UTRs.

Primer pairs that yielded positive alternative splicing results were: F21/R11, F29/R11, F6/R2', and F32/R12 (see **Figures 6 and 11** for primer locations). Primer pair F21/R11 should amplify exons 1 through 10, however when this PCR was performed, multiple, smaller bands were seen in addition to the predicted product. These bands were cloned into pGEM-Teasy and individual clones were sequenced. It was determined that the smaller products were legitimate, but lacked one or more of the exons (**Figure 11**). In total 6 smaller products were obtained, with the following exon composition:

- 1,2,3,4,5,6,7,9,10
- 1,2,3,8,9,10
- 1,2,3,9,10
- 1,2,9,10
- 1,8,9,10
- 1,9,10

These splice variants all maintain the appropriate reading frame with the exception of the exon 1,2,9,10 variant. The F29/R11 primer pair showed that exon 8 is either retained or spliced in all tissues tested with no preference for either variant. The different alternative splicing of exons 2-8 was observed with all cDNA synthesis primers tested (R30, R26, R11, R6 see **Figure 6, 11** for primer location), suggesting these splice variants are not specific to any particular *CECR2* transcript.

Two other primer pairs identified alternative splicing events that removed only a portion of an exon. Primer pair F6/R2' identifies a splice variant that removes 84 bp from the 3' end of exon 14, while primer pair F32/R12 identified a splice variant that removes

165 bp from the 3' end of exon 16 (**Figure 11**). All the human alternative splice variants (see above) were also identified in mouse (with the exception of the exon 2 to 9 splice, which does not conserve the reading frame) suggesting that they are biologically relevant. Assistance on the mouse RT-PCR experiments was provided by technologist Melanie Kardel. It should be noted, however, that RT-PCR is not quantitative so we were unable to determine the relative frequency of each of the alternative transcripts. Based on Northern blots, the 10.0 and 9.5 kb transcripts appear to be the most prevalent (**Figure 10**). Western blots with Cecr2 specific antibodies to exon 17, show a band of approximately 160 kDa (T. Ames, personal communication) in adult liver, which suggests that the largest version of *Cecr2* which includes all the coding exons (exons 1-19{or 19' or 19"}), is the most prevalent in that tissue. Adult mouse brain westerns show a transcript of approximately 140 kDa, the Cecr2 exon composition of which is unclear at this time (T. Ames, personal communication).

The combination of alternative splicing and alternative polyadenylation generates a large number of possible transcripts from the *CECR2/Cecr2* locus. A subset of the transcripts are shown in **Figure 11**. I was unable to determine if certain alternative splices are found only in combination with certain UTRs, or are restricted to certain tissues, and as such I have been unable to lower the number of putative transcripts from the current theoretical number of 78.

3.8 Prediction of *Cecr2* genomic structure and mRNA sequence

The mouse copy of *Cecr2* was identified based on GENSCAN gene prediction of BACs 555D9 (*AC009192*) and 67D14 (*AC006447*) and was confirmed by homology

mapping to the human cDNA sequence. *Cecr2* retains the same overall genomic structure as the human copy with minor variations (**Table 2**). ESTs confirm three distinct polyadenylation sites within exon 19, which corresponds to the human exons 19, 19' and 19". To date, the only alternative transcripts not identified in mouse are the alternative exon 13', which results in the short, 13 exon transcript *CECR2b* and the alternative splice between exons 2 and 9. Considering that the exon 2-9 splice in humans puts *CECR2* out of frame, it is not surprising that it is not found in mouse, suggesting the human variant is a splicing artifact.

3.9 Sequence analysis of *CECR2/Cecr2*

Comparison of the human amino acid sequences of CECR2 and Cecr2 showed that the amino 1/3 of the protein was more highly conserved than the distal 2/3. Over the first 540 amino acids, which contains all the recognizable amino acid motifs (see below), CECR2 and Cecr2 are 94% identical and 96% similar. However, over the remaining ~950 aa the two proteins are only 74% identical and 79% similar (**Figure 12**).

The putative amino acid sequences of CECR2/Cecr2 were subjected to various computer searches in an attempt to learn more about possible function. CECR2 was predicted to be nuclear by PSORT II (Nakai and Horton 1999). Motif searches of CECR2 with the MOTIF webserver (http://motif.genome.ad.jp/) against the Pfam (Bateman *et al.* 2002), Prints (Attwood *et al.* 2002), Blocks (Henikoff *et al.* 1999) and Prosite (Bucher *et al.* 1994; Falquet *et al.* 2002) databases identified several motifs that added support to the nuclear localization prediction. Motifs predicted included a bromodomain, which is a motif that has been shown to bind to acetylated histones (Dhalluin *et al.* 1999) and is

found commonly in transcriptional coactivator proteins such as CBP/p300 (Bannister and Kouzarides 1996; Ogryzko *et al.* 1996), P/CAF (Yang *et al.* 1996) and GCN5 (Brownell *et al.* 1996). Other motifs identified include an AT hook, which is a motif found within the high mobility group of proteins (HMG) and has been shown to be a DNA binding motif with a propensity for binding AT rich DNA sequences in the minor groove of DNA (Reeves and Nissen 1990). A putative ATP/GTP binding motif was also identified, but the significance of this prediction is unclear. However, the AT Hook and bromodomain motifs are consistent with a nuclear protein that associates with DNA. Locations of these motifs are shown in **Figure 13**. In addition several LxxLL-like motifs were identified within CECR2. These motifs, and variants thereof, have been shown to be protein interaction motifs found in transcriptional coactivators (Heery *et al.* 1997; McInerney *et al.* 1998; Webb *et al.* 2000; He *et al.* 2002). One of each of the following variants were identified: LxxLL (exon 7), IxxLL (exon 2), IxxLI (exon 2), and FxxLF (exon 1). Each of these motifs was found in the corresponding location in *Cecr2*. Basic BLAST searches against the non-redundant protein database only yielded significant homology to other proteins, excluding *Cecr2* and *Drosophila CG10115*, within a region in the middle of CECR2, which contains the bromodomain motif.

The 3' UTR (exon 19, 19' and 19") of *CECR2/Cecr2* was of interest because of the disparity in the size of exon 19" from 19 and 19', and the fact that the 3 distinct polyadenylation locations were found to be conserved between human and mouse. Because of this, a more detailed sequence analysis was undertaken. The 3' UTR sequences of *CECR2* and *Cecr2* were compared by BLAST, and three regions of homology were identified (**Figure 14**). Region 1 spans ~260 bp and is 85% identical,

region 2 spans ~500 bp and is 82% identical, and region 3 is ~215 bp and is 90% identical. Region 1 is close to the 5' end of exon 19, regions 2 spans the polyadenylation signals utilized by exons 19 and 19', while region 3 covers the most 3' sequence of exon 19", which presumably contains a polyadenylation signal. However, no canonical polyadenylation signal was identified near the 3' end of exon 19". The UTRs were checked for repeats by Repeatmasker, and it was determined that *CECR2* has an AluSx inserted in the distal region of exon 19", which is not found in *Cecr2*. It was also observed that *CECR2* contains a low complexity repeat $(CCCUU)_n$ between the Alu and the 3' end of exon 19". This repeat was not found in *Cecr2*. Similar (CU) rich repeats in 3' UTR regions have been previously implicated in regulation of translation (Collier *et al.* 1998), mRNA stability (Yeap *et al.* 2002) and nuclear/cytoplasmic transport (Fan and Steitz 1998a; Fan and Steitz 1998b) via interaction with the RNA binding proteins HuR, CP1 and CP2.

3.10 Expression of CECR2 in insect cells

Several proteins that contain bromodomains also exhibit histone acetyltransferase (HAT) activity. Unfortunately there is no sequence consensus for HAT domains, so it must be determined experimentally. In order to test for histone acetyltransferase activity in CECR2, a full length version of the *CECR2* open reading frame was subcloned into a modified version of the insect expression vector pMIB/V5-HIS-A, in addition to two halves of the protein in separate constructs. The transcriptional coactivator *GCN5* was also cloned into the same vector as a positive control, as it had been previously shown to have HAT activity (Brownell *et al.* 1996). The constructs were tested for protein

expression after 2 days of transient expression, and it was determined that all constructs were expressing protein, but it was not being secreted into the media (**Figure 15**). After it was determined that the cells were expressing the constructs they were placed under Blasticidin selection in order to force the integration of the expression vectors into the host genome. These cells were then tested for expression (**Figure 16**). As with the transiently expressing cells, the cell lines with the integrated expression vectors did not secrete the expressed proteins into the media (data not shown). Full length CECR2 and GCN5 were subsequently used in a histone acetyltransferase assay.

3.11 Histone acetyltransferase assay of CECR2

The HAT assay involves a reaction in which ^{3}H-aectyl coenzyme A, calf thymus histones, buffer and the recombinant protein to be tested are mixed and incubated. The resulting mixture is resolved on a SDS PAGE gel, which is then stained and fluorographed. Should CECR2 have HAT activity, the radiolabel should be moved from the acetyl CoA to one of the four core histones. I was never able to successfully show HAT activity for either CECR2 or GCN5, which is known to acetylate Histone H3 preferentially (Herrera *et al.* 1997). There was always a background level of signal regardless as to whether the cell line was expressing CECR2, GCN5 or an unrelated protein that did not possess HAT activity (data not shown). This is likely due to technical issues in the protein isolation and the assay itself. Due to time constraints no further experiments were attempted beyond these initial experiments.

3.12 Creation/characterization of CECR2 transgenic mice

In an attempt to determine if overexpression of CECR2 in mice results in abnormal development, transgenic mice were generated with the BAC 357F7 transgene, by pronuclear injection (as shown in **Figure 18a**). Dr. Peter Dickie (HSLAS) performed pronuclear injection of circular BAC DNA into mouse FVB/n ooctyes, which were then implanted into pseudopregnant females. From a single round of approximately 100 injections, we obtained 10 pups, of which 2 clearly contained the transgene (**Figure 17a**). Of the 10 founders (designated Lines 1-10), 1 and 10 went on to produce progeny that carried the transgene. Initial Southern blot data **(Figure17a)** suggested that Line 10 had a higher transgene copy number than Line 1, but we didn't determine an absolute copy number as we soon shifted to PCR genotyping of progeny (**Figure 17b**).

Pups were genotyped by PCR of tail DNA **(Figure 17b)** with human BAC 357F7 specific primers (*CECR2* F26/R28). The primer set was in exon 19" of *CECR2* and was designed to amplify only the human copy under the reaction conditions. Initially, transgenic animals were mated to normal FVB/n animals. Line 1 produced large litters (~10 pups) every 3-4 weeks, when females and males were left caged together. Subsequently we mated Line 1 transgenic animals together in an attempt to increase the copy number of the transgene. Unfortunately, we had no method to determine whether animals were heterozygous or homozygous for the transgene. However, we saw no change in the size or regularity of the litter in Line 1 even when two transgenic animals were mated together. No abnormal phenotypes were observed in these animals.

In contrast to Line 1, Line 10 was problematic from the start. Only four pups were produced from the founder animal. It proved to be very difficult to produce another litter

from these pups, as several litters were lost *in utero*. Eventually the line was propagated to F4, but litters were generally small, and mothers tended to appear pregnant and then fail to give birth. Due to fear of losing Line 10, we did not sacrifice pregnant females to observe the embryos. However, the transgenic animals that were propagated showed no abnormalities. The eyes of both Line 1 and 10 animals were examined by an ophthalmologist, and no abnormalities were observed. Preliminary RT-PCR data with *CECR2* specific primers suggested that *CECR2* was indeed expressed in both lines (data not shown), although a more detailed analysis is required. Positive RT-PCR results suggested that the transgene was intact and producing mRNA transcripts. However, we did not determine absolute copy numbers of the transgene by Southern blot, or relative copy numbers by quantitative RT-PCR. The integration sites of the transgenes are currently unknown, but FISH of mouse chromosomes with a BAC 357F7 probe will be attempted in the future to address this question.

Both Line 1 and 10 transgenics were subsequently crossed to *Cecr2* knockouts (see **3.13**) to determine whether or not the transgenes would rescue the mutant phenotype, and were therefore producing functional protein, and also to ensure that the transgenes were not lost if the FVB/n transgenic lines were lost (this is explained in more detail in section **3.17**).

3.13 Creation of *Cecr2* "knockout" mice via an ES cell genetrap

During the initial BLAST analysis of *CECR2/Cecr2*, a 5' RACE product from a mouse genetrap clone was identified (data not shown). The 5' RACE sequence of the genetrap clone (CT45) matched sequence from *Cecr2* exon 7 suggesting that the genetrap

clone had inserted in intron 7. As we were investigating the possibility of dosage sensitivity of *CECR2/Cecr2*, we decided to obtain the ES cell line containing the genetrap and generate a mouse that contained a *Cecr2* mutation. A male ES cell line (CT45) from mouse strain 129P2/Ola, which contained the pGT1 genetrap (Wilson *et al.* 1995) within the *Cecr2* locus, was obtained from Dr. Wendy Bickmore (Tate *et al.* 1998). The ES cell line was grown by Dr. Roseline Godbout and used for blastocyst injection by Dr. Peter Dickie into blastocysts from mouse strain C57BL/6J (as shown in **Figure 18b**). These injected blastocysts were implanted into pseudopregnant females. Strain 129P2/Ola has a yellow-agouti coat colour and strain C57BL/6J has a black coat colour. Therefore, any chimeras produced should be multicoloured. We obtained 3 chimeras from the blastocyst injections: one female, and two males. As the ES cell line was male in origin the female chimera was not mated. The two male chimeras were originally mated to C57BL/6J animals but we were unsuccessful in obtaining progeny. Subsequently the two chimeras were mated to BALB/c (albino) females, and one of the two proved to be fertile (**Figure 19a**). The one fertile chimera produced progeny from the 129P2/Ola ES cell derived background at a rate of approximately 8%. When mated to BALB/c females, progeny from the ES lineage were chinchilla (slate grey) in colour while progeny from the C57BL/6J background were agouti (**Figure 19b**). Of the chinchilla F1 animals, only half would be expected to contain the genetrap insertion as the ES cell line was heterozygous for the insertion.

To facilitate genotyping of these animals, the exact position of the insertion was identified via PCR between *Cecr2* intron 7 and the end of the genetrap vector, pGT1 (data not shown). This PCR product was sequenced to identify the precise insertion site.

This allowed primers to be designed for PCR genotyping that allowed differentiation between +/+, +/- and -/- animals (**Figure 20a**). A multiplex PCR reaction was employed that contained 2 primers from *Cecr2* intron 7 spanning the insertion site, one primer in the genetrap and 2 primers from the *Sry* gene to facilitate sexing (**Figure 20b**). Genetrap positive progeny ($Cecr2^{+/pGT1}$) were identified by PCR and were initially crossed to BALB/c females to generate more heterozygotes. Subsequently, $Cecr2^{+/pGT1}$ animals were intercrossed to generate homozygotes. It quickly became clear that there was a lethality associated with the homozygous mutant genotype, as only 4 $Cecr2^{pGT1/pGT1}$ animals were identified by genotyping of 100 weaned pups resulting from *Cecr2* heterozygote intercrosses. In total, only 7% of the weaned progeny were genotyped as $Cecr2^{pGT1/pGT1}$ (**Table 3**), suggesting some *in utero* or perinatal lethality. Subsequently, pregnant mothers were sacrificed in order to analyze embryos *in utero*. Embryos from 9.5 dpc to 19.5 dpc were observed with the neural tube defect exencephaly (**Figure 21a, Figure 24**), while 19.5 dpc embryos appeared to be lacking eyelids (**Figure 21a**). At this point it was observed that the genetrap mutation appeared to segregate with the exencephaly phentoype when in the homozygous state and also at low frequency when in the heterozygous state. Exencephaly occurred in 67% of the $Cecr2^{pGT1/pGT1}$ embryos and 3.5% of $Cecr2^{+/pGT1}$ embryos on the 129P2:BALB/c mixed background (**Table 3**). Exencephaly results from failure of closure of the anterior neural tube during neurulation between 8.5 and 9.5dpc. At this point expression analysis of *Cecr2* (by way of the *LacZ* reporter gene) was performed to determine if the expression profile was such that loss of *Cecr2* could plausibly perturbate neurulation (see **3.14**).

3.14 RT-PCR of *Cecr2* in *Cecr2*$^{pGT1/pGT1}$ mice

As the pGT1 vector is a splice trap vector and does not involve homologous recombination/replacement of the *Cecr2* locus, we wanted to determine if our *Cecr2*$^{pGT1/pGT1}$ mice were producing any normal *Cecr2* transcript, or were null for the allele. *Cecr2* specific primers located 3' to the genetrap (intron 7) were used to perform RT-PCR on RNA obtained from 14.5 dpc *Cecr2*$^{pGT1/pGT1}$ embryos. Three separate primer sets were used and each set produced the expected product (**Figure 22**), showing that the pGT1 allele of *Cecr2* was not null. However, without use of quantitative RT-PCR we were unable to determine whether the expression level of *Cecr2* was affected in the mutants. Northern blots were attempted with probes corresponding to the RT-PCR products in Figure 21, but did not yield interpretable signals. This was not a surprising result following the failure of mouse embryo RNA *in situ* hybridizations.

Once antibodies are produced against Cecr2, we will be able to determine if wildtype Cecr2 is being produced in *Cecr2*$^{pGT1/pGT1}$ mice, and to what extent. From the *Cecr2* reporter gene studies it is clear that the pGT1 genetrap does infact contribute to *Cecr2* mutant transcripts, as β-galactosidase enzyme is produced (see below). Further study is required to determine the extent of alternative splicing around the genetrap. This is of interest due to the fact that intron 7 is alternatively spliced in *CECR2/Cecr2*. Study of splicing around the genetrap is ongoing.

3.15 Reporter gene expression of Cecr2 in mouse embryos

Previous attempts at RNA *in situ* hybridization on mouse embryos had proved problematic, so Xgal staining of 10.5-14.5 dpc *Cecr2*$^{+/pGT1}$ embryos was performed.

From this experiment it was determined that β-Gal expression is most concentrated in the central nervous system, specifically the forebrain, eyes, spinal column and spinal ganglia at 13.5 dpc (**Figure 23**). Expression was also seen in the nasal cavity and the mesenchyme of the developing limbs. There did appear to be a basal level of expression found in many other tissues, but never in the heart. It was also noted that general expression patterns were similar in 10.5 dpc embryos (**Figure 25**) to that of 13.5 dpc embryos. Due to the appearance of staining deeper inside the embryo, embryos were cleared in order to better visualize interior regions (**Figure 26**). Looking at the cleared embryos, it became apparent that staining was present in the intercostal region, although from the whole mount embryo it was not clear if staining was in the rib cartilage, intercostal muscles, or elsewhere. Also observed in the cleared embryos were two large ganglia in the lateral midbrain region. Several Xgal stained *Cecr2*$^{+/pGT1}$ and *Cecr2*$^{pGT1/pGT1}$ embryos were paraffin embedded and sectioned by microtome. From this analysis, it became clear that the β-Gal expression in the body wall was not in the rib cartilage, but in the intercostal region, likely in the muscle (**Figure 27**). Some other weak β-Gal expression was noted in the developing kidneys and thyroid primordium (data not shown).

Due to the *Cecr2* reporter gene expression in the intercostal region and the developing limbs, a skeletal analysis of *Cecr2* embryos was performed to determine if there were any skeletal abnormalities in *Cecr2* embryos (see **3.15**).

3.16 Skeletal analysis of Cecr2 mutant embryos

Cecr2 embryos of all three genotypes were subjected to skeletal analysis to determine if the *Cecr2* disruption was causing any skeletal abnormalities. Embryos were stained with a cartilage specific dye, Alcian Blue 8GX, prior to genotyping to avoid bias in analysis. Looking at the stained embryos it became clear that the only abnormality observed was the lack of a forming cranium in the $Cecr2^{pGT1/pGT1}$ mice, which were penetrant for exencephaly (**Figure 28**). No abnormalities were observed in the ribs or limbs where *Cecr2* reporter gene expression was found nearby. The lack of skeletal development in the cranial region is likely a secondary effect of the exencephaly phenotype as lack of skull development is always found in exencephalic embryos

3.17 Folic acid rescue experiment in Cecr2 mutant mice

Due to the observation that some neural tube defects can be rescued by maternal supplementation of folic acid we decided to ask if *Cecr2* mutants could be rescued by folic acid. Previously reported methods of folic acid supplementation include dietary supplementation (*Cd*; Carter *et al.* 1999), intraperitoneal injection (*Cart1*; Zhao et al. 1996), or embryo culture (*Pax3*; Fleming and Copp 1996). All three methods have been used successfully with intraperintoneal being the most common method used.

We used the intraperitoneal injection of folic acid as our delivery method. *Cecr2* heterozygotes were intercrossed, and starting the day of mating, females were injected interperitoneally with 10 mg/kg of folic acid (or PBS as a control) until day 12 or 13 when the females were euthanised and the embryos analyzed for NTD and genotyped. As shown in **Table 4**, it is clear that folic acid does not appear to rescue the *Cecr2* mutation,

even with the limited sample size. The rates of exencephaly are not reduced from the PBS associated rate. The penetrances of 55% and 80% are estimates based on the small sample sizes observed. A more accurate rate is that found in **Table 3** (around 67%), which is based on a much larger sample size for the 129P2:BALB/c mixed background. Regardless, it does not appear that folic acid has any influence on the rates of exencephaly in *Cecr2*$^{pGT1/pGT1}$ mice, with the caveat that this is a rather small sample size. Other studies have typically reported results using between 20 and 30 homozygous animals (Zhao et al. 1996; Fleming and Copp 1996) athough sample sizes as small as 18 homozygous animals have been reported (Barbera *et al.* 2002). We decided not to further test folic acid on *Cecr2* mutant animals while they were on the mixed 129:BALB/c background due to the inherent genetic heterogeneity of this background. We will repeat this experiment once the *Cecr2* pGT1 allele is moved to 129P2 and BALB/c purebreeding lines.

3.18 *CECR2* transgene complementation experiment in *Cecr2*$^{pGT1/pGT1}$ mice

In order to determine if the human *CECR2* transgene was actively producing protein and that this CECR2 protein could functionally complement a *Cecr2* mutant, the human transgene was crossed into the *Cecr2* knockout background (**Figure 29**). The *CECR2* transgene contained within BAC 357F7 was in a pure FVB/n background, while the *Cecr2* knockout was in a mixed 129P2:BALB/c mixed background. This meant that the exencephaly penetrance data would have to be reassessed for this new background (129P2:BALB/c:FVB/n). A (FVB/n) *CECR2* transgenic male was mated to (129P2:BALB/c) *Cecr2*$^{+/pGT1}$ females to generate *Cecr2*$^{+/pGT1}$, +TgN(*CECR2*) males that

were used as stud mice. These males were subsequently mated back to (129P2:BALB/c) *Cecr2*$^{+/pGT1}$ females to generate all genotypes of *Cecr2* with and without the *CECR2* transgene. The major focus of this rescue experiment, due to resources, was the Line 10 transgene (described in **3.12**). From the initial litter of stud males, two *Cecr2*$^{+/pGT1}$,+TgN(*CECR2*) males were used for the experiment and one *Cecr2*$^{+/pGT1}$ male was used as a control to recalculate the exencephaly penetrance.

From the control animals it became clear that the presence of the FVB/n background muted the exencephaly penetrance. Penetrance was calculated at 47% 129P2:BALB/c:FVB/n, as compared to 67% in the 129P2:BALB/c background. In the experimental animals, it did not appear that the presence of the *CECR2* transgene reduced the rates of exencephaly in either sub-line of Line 10 (**Table 5**). Progeny of Line 10 ♂#17 had an exencephaly penetrance of 20% regardless of whether the transgene was present or absent. Progeny of Line 10 ♂#25 showed similar results to that of Line 10 ♂#17. Penetrance was 22% in *Cecr2*$^{pGT1/pGT1}$,+TgN(*CECR2*) as compared to 25% when the transgene was absent. The penetrance rates do vary somewhat between the progeny of the experimental males and the control male. This is likely due either to sample size or differences in modifier loci between the stud males. Due to time constraints, Line 1 rescue was not tested beyond the preliminary stage, and as such will require further experimentation to determine if it will complement the *Cecr2* mutation.

The results from the transgene rescue experiment suggest that the *CECR2* transgene is not rescuing the *Cecr2* mutation, but reasons for this are currently unclear. We do not know if the transgene is generating functional CECR2 protein, and whether or not the human protein is capable of function in mouse.

3.19 Excision of a P element insertion within the *Drosophila melanogaster CECR2* (*CG10115*)

BLAST analysis of the putative protein sequence of *CECR2* yielded 31% identity and 52% similarity over the first 350 aa to the *Drosophila melanogaster* hypothetical protein *CG10115*, in the same region of *CG10115*. This protein is based on gene prediction from the *Drosophila* genome sequencing/annotation project. The bromodomain is conserved in *CG10115*, suggesting this may be an ortholog of *CECR2*. In collaboration with Dr. John Locke, we attempted to generate a deletion within *CG10115*. A search of Flybase (www.flybase.org) using the genomic region containing *CG10115* identified a strain with a P-element insertion within the *CG10115* locus. As *Drosophila* is readily amenable to genetic manipulation, we decided to attempt to create a deletion within *CG10115* by excising the P-element. This would allow us to observe any potential phenotype from a disruption of *CG10115* in *Drosophila*, and if one was found would allow for a future genetic suppressor screen to identify potential genetically interacting proteins. The P-element containing strain (13156) was first tested to determine if in fact the P-element was in the stated position (*CG10115* intron 1).

The P element location and orientation was confirmed by PCR between the P-element and the intron sequence surrounding it. P element specific primers SK08 and 11rpt5 were used in combination with primers from *CG10115* intron 1 (Dm CECR2 F4 and R4) (**Figure 30a**). The results of these PCR reactions confirmed that the P element was indeed within intron 1 of *CG10115*, and was in an inverted orientation with respect

to the *CG10115* transcript (**Figure 30a,b**). This is likely why it does not appear to disrupt the *CG10115* locus (as the strain is viable with the P-element in a homozygous state).

In collaboration with Dr. John Locke, we created a variety of lines in which the P element was excised in an attempt to create a small deletion, thus disrupting *CG10115*. A three-step crossing scheme (**Figure 31**) was performed by Dr. Locke to produce individual males, each containing a unique excision event of the P element. Briefly, Δ2-3 recombinase was crossed into the P-element background so that excision could take place. Then the male progeny were crossed to a strain carrying a deficiency for the region surrounding *CG10115*. This produced the excised P element allele over a deficiency for the region. If *CG10115* was disrupted by the P-element excision, it would effectively be a homozygous mutation. Progeny of the third cross were scored for lethals in the Ubx^+, Sb^+ and Hu^+ (or Ly^+) category. All flies generated in these categories were viable.

I performed single fly PCR on the P-element excision progeny to determine if a deletion had been generated. In addition to a positive PCR control (Su(var)3-7), the primer set Dm CECR2 F4/R4, which spans the P element insertion site, was used to determine if a deletion was generated. A negative result for F4/R4 in combination with a positive result from Su(var) 3-7 Fwd/Rev would indicate a deletion was generated. If a deletion was detected, other nearby primer sets would be tested to determine the extent of the deletion. I tested 49 individual flies (excision events) from two separate experiments, but did not find any lines in which a deletion was generated (**Figure 30c**). The project was not carried beyond this point.

Table 2. Comparison of exon/intron sizes between *CECR2* and *Cecr2*.

	Exon size (bp)		Intron size (bp)	
	CECR2	*Cecr2*	*CECR2*	*Cecr2*
1	~544	~506	106779	54187
2	95	95	19710	10291
3	184	184	1823	2307
4	140	140	1081	682
5	105	105	2346	2412
6	50	50	1715	1006
7	170	170	6796	5622
8	84	84	12215	6987
9	154	154	13494	1635
10	130	130	1287	1756
11	38	38	82	86
12	92	92	261	776
13	127/216	127	1292	904
14	305/389	305/389	1122/1038	711/627
15	129	126	189	123
16	689/847	652/817	5310/5145	2658/2493
17	1417	1417	2461	2719
18	112	112	707	1030
19	1111/1307/5313	1054/1259/4775		

Table 3. *Cecr2* Genotype/phenotype data from the 129P2:BALB/c mixed background. In the genotype totals category the percentage refers to the percentage of the animals that were of that particular genotype.

	pGT1/pGT1	*+/pGT1*	*+/+*
Weaned			
Crooked Tail	5	2	0
Normal	26	256	148
Penetrance	16%	0.8%	0%
Genotype totals	31 (7%)	258 (59%)	148 (34%)
Embryos			
Exencephaly	24	4	0
Normal	12	113	38
Penetrance	67%	4%	0%
Genotype totals	36 (19%)	117 (61%)	38 (20%)

Table 4. Results of the maternal folic acid supplementation experiment. *Cecr2* genotypes and phenotypes are based on analysis of 12-14dpc embryos on a 129P2:BALB/c background.

	pGT1/pGT1	*+/pGT1*	*+/+*
Folic Acid			
Exencephaly	8	0	0
Normal	2	30	13
Total	10	30	13
Penetrance	80%	0%	0%
PBS			
Exencephaly	6	0	0
Normal	5	8	7
Total	11	8	7
Penetrance	55%	0%	0%

Table 5. Results of the *CECR2* transgene rescue experiment. TgN refers to the BAC 357F7 transgene containing *CECR2*. Data represent *Cecr2* genotype/phenotype data from the embryonic progeny of the lines listed. The control data is to determine penetrance on the 129P2:BALB/c:FVB/n mixed background.

	pGT1/pGT1	*pGT1/pGT1*, TgN	*+/pGT1*	*+/pGT1*, TgN	*+/+*	*+/+*, TgN
Line 10 ♂#17						
Exencephaly	5	3	2	0	1	0
Normal	20	12	40	64	21	27
Total	25	15	42	64	21	27
Penetrance	20%	20%	3%	0%	5%	0%
Line 10 ♂#25						
Exencephaly	3	4	0	1	0	0
Normal	9	14	42	53	15	29
Total	12	18	42	53	15	29
Penetrance	25%	22%	0%	2%	0%	0%
Control						
Exencephaly	10	N/A	5	N/A	0	N/A
Normal	11	N/A	65	N/A	28	N/A
Total	21	N/A	70	N/A	28	N/A
Penetrance	46%	N/A	7%	N/A	0%	N/A

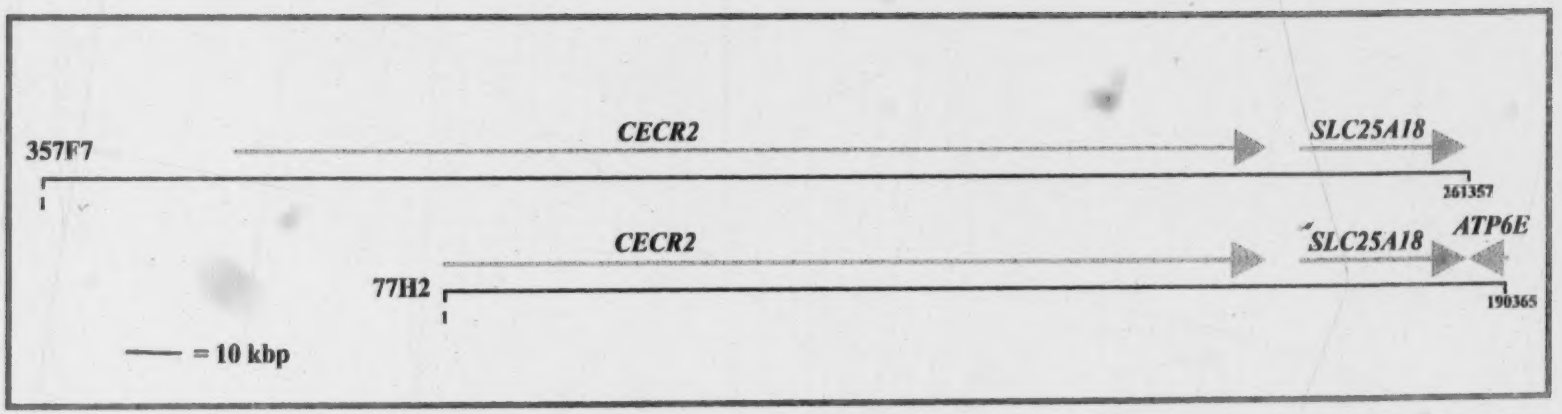


Figure 3. Schematic showing the overlapping BAC clones 357F7 (top) and 77H2 (bottom) and their respective gene content. Genes shown in blue are complete, while genes in red are incomplete. Direction of gene transcription is noted with an arrow.

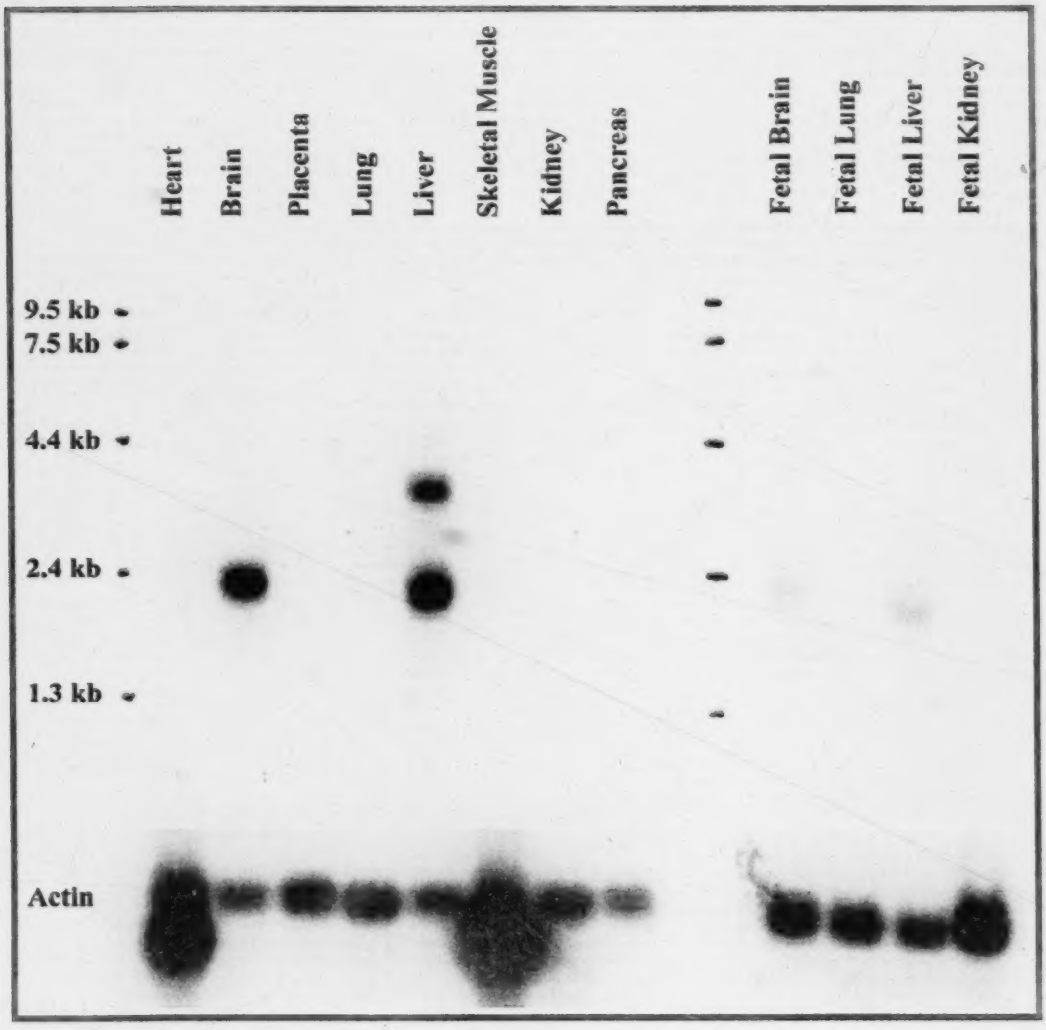


Figure 4. Expression profile of *SLC25A18*. Clontech™ human MTN Northern blots were hybridized with EST 28949. Actin loading controls are below each blot.

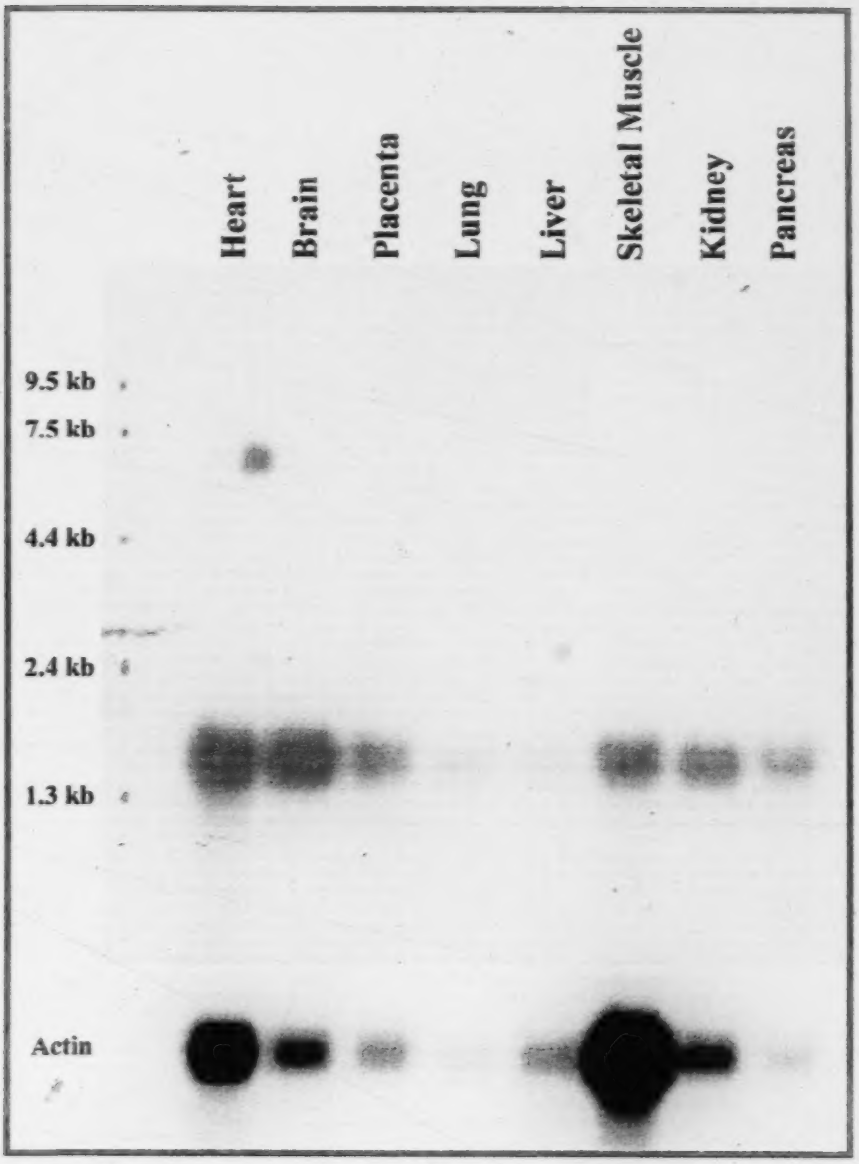


Figure 5. Expression profile of *ATP6E*. A Clontech™ adult human MTN Northern blot was hybridized with the 61EW probe. Actin loading control is shown below the blot.

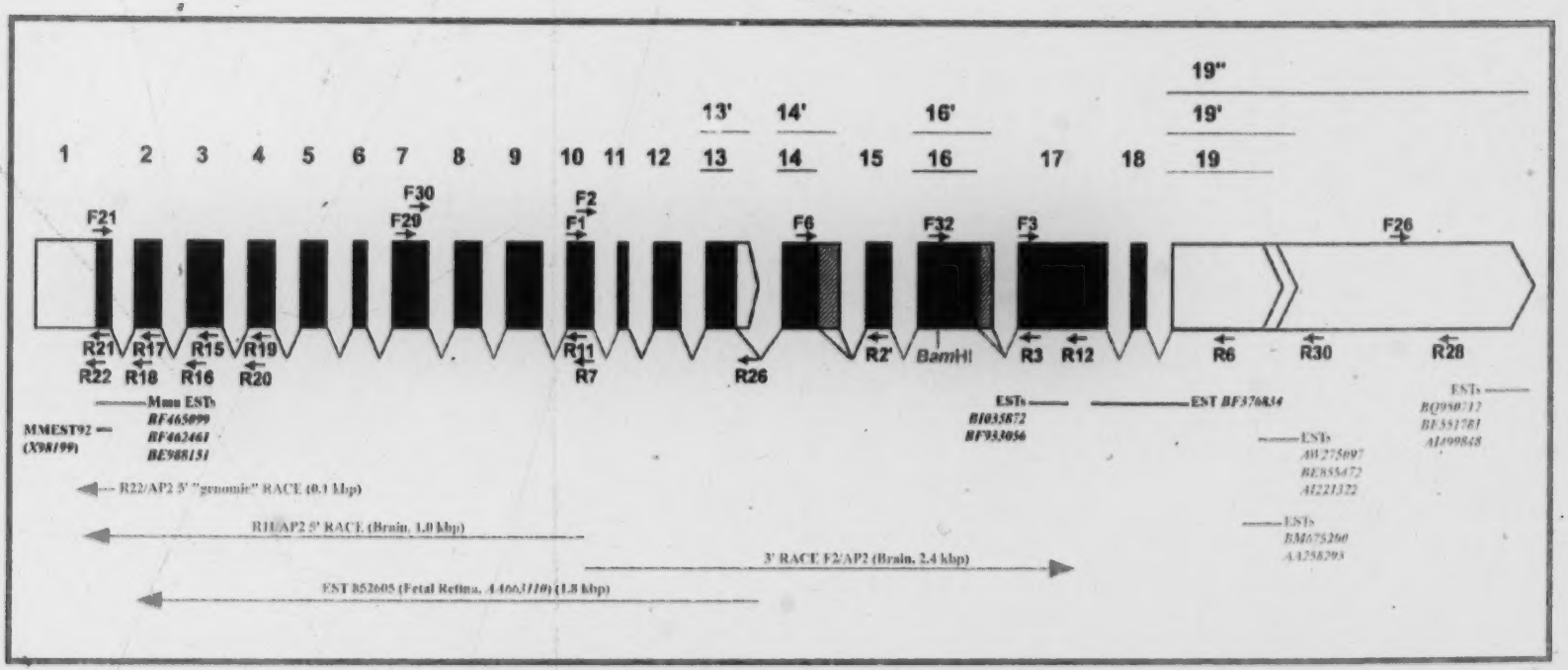


Figure 6. Schematic of *CECR2* (not to scale) showing the location of various ESTs, RACE products and primers. The mouse (Mmu) ESTs that span exons 1 and 2 only have significant identity over the region of the *CECR2* open reading frame (shown in black). The gray bar represents a region of lower identity between the 5' UTRs of *CECR2/Cecr2*.

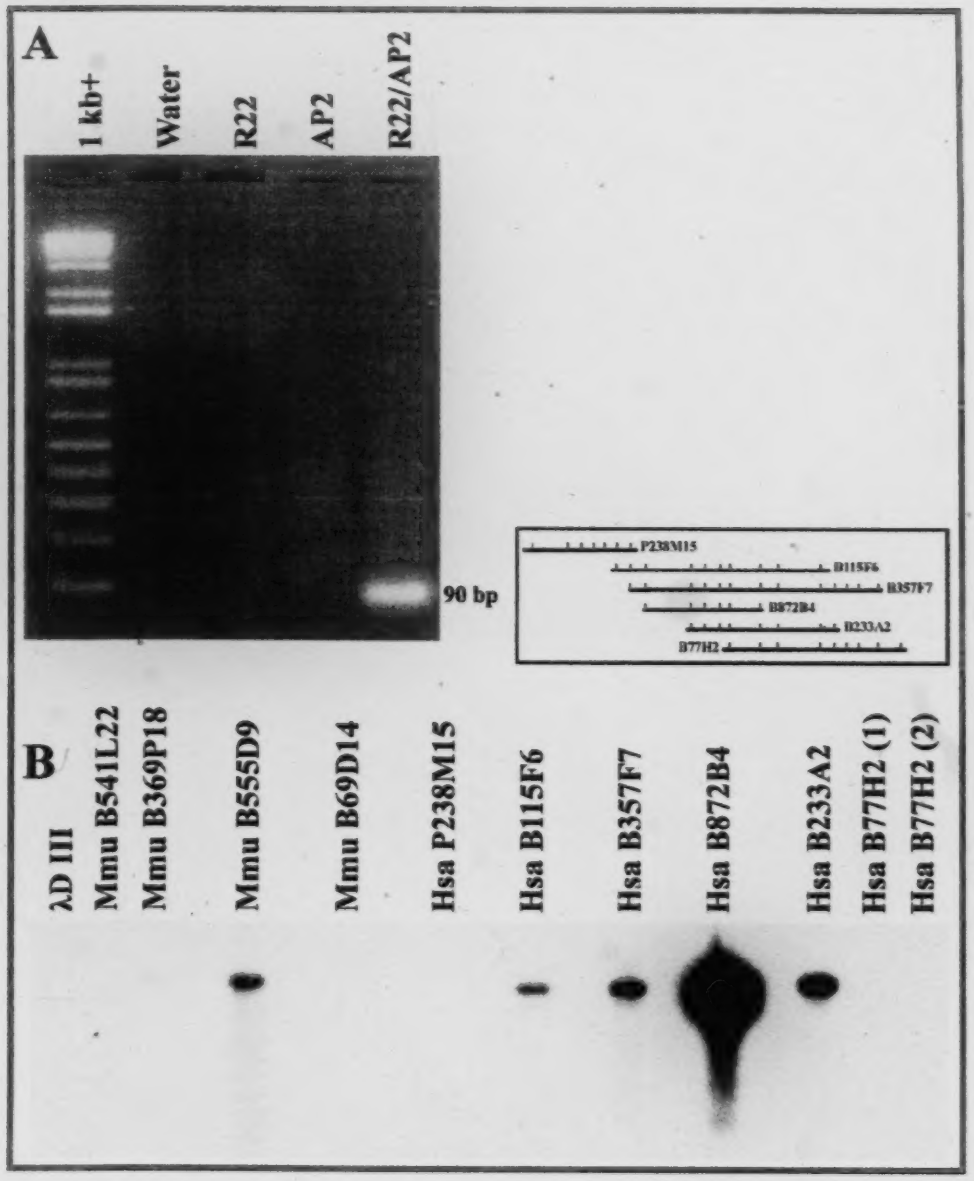


Figure 7. **(A)** Genomic "RACE" of the region surrounding *CECR2* exon 1. The PCR library was generated from *Pvu* II cut genomic DNA ligated to linkers. A product of approximately 90 bp is observable in the R22/AP2 lane, but not in the 3 control lanes. **(B)** Southern blot of human and mouse CESCR BACs/PACs using the R22/AP2 RACE product from (A). *CECR2* exon 1 is present on several overlapping human clones and the homologous clones in mouse. Note however, the absence of signal from human BAC 77H2 (human clones are shown above in box, mouse clones in **Figure 2**).

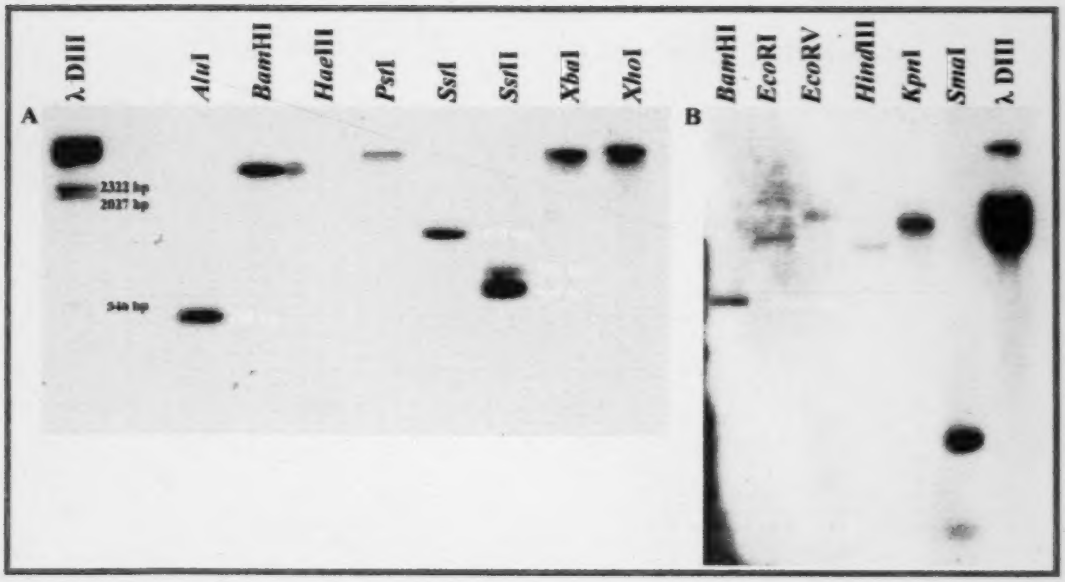


Figure 8. Southern blots of BAC 357F7 cut with various enzymes. Both blots were probed with a *CECR2* exon 1 probe (R22/AP2). The *Sst*I and *Bam*HI fragments were subsequently subcloned and sequenced in order to determine the extent of the missing sequence from the Genbank version of BAC 357F7.

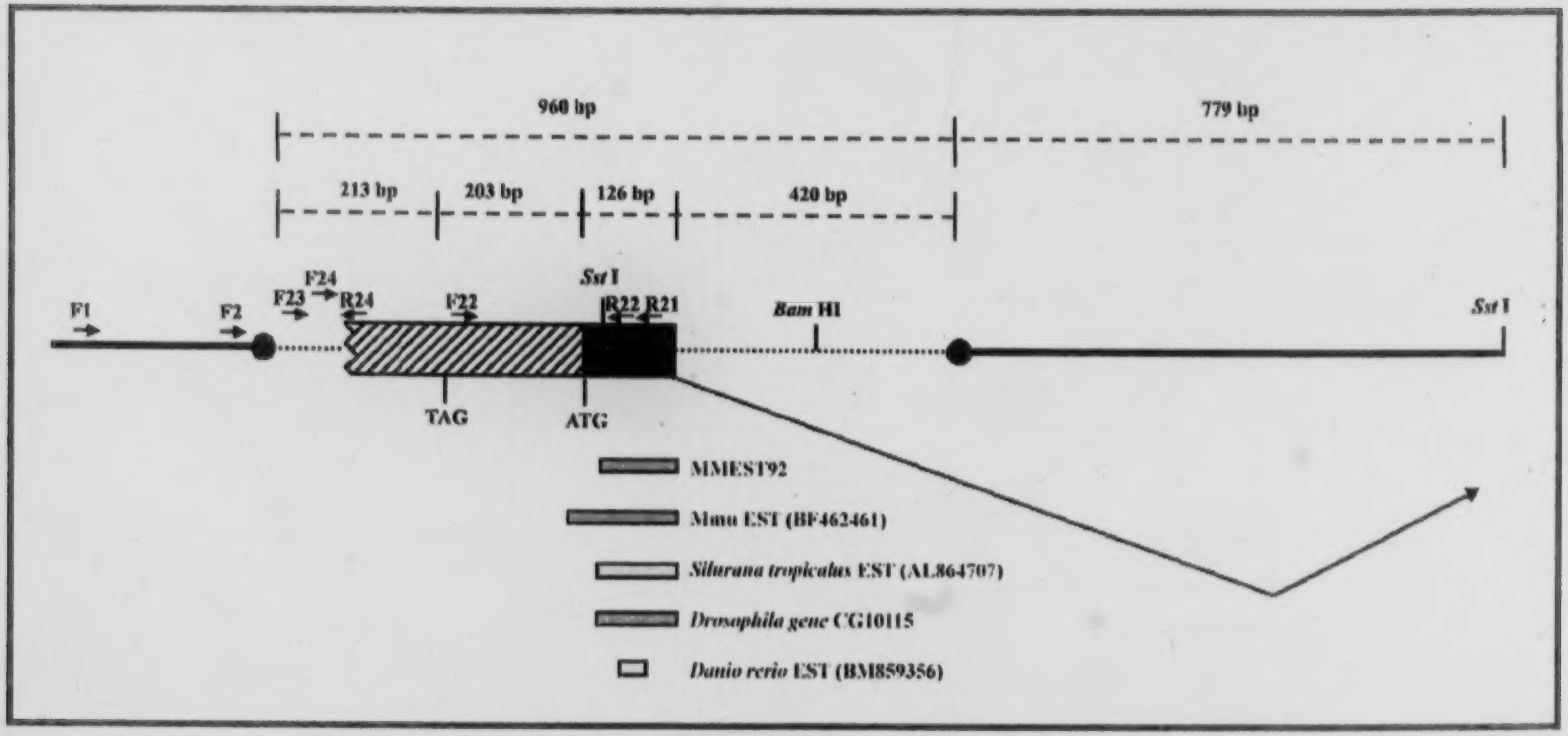


Figure 9. Schematic depicting the deletion area surrounding *CECR2* exon 1 in BAC 357F7. The area between the black circles represents the area missing from the Genbank version of BAC 357F7. The solid box represents the coding region of *CECR2* exon 1, and the hatched box represents the putative 5' UTR. Primer locations are noted, as are regions of similarity to various ESTs from other species (DNA and protein from mouse, protein only for others).

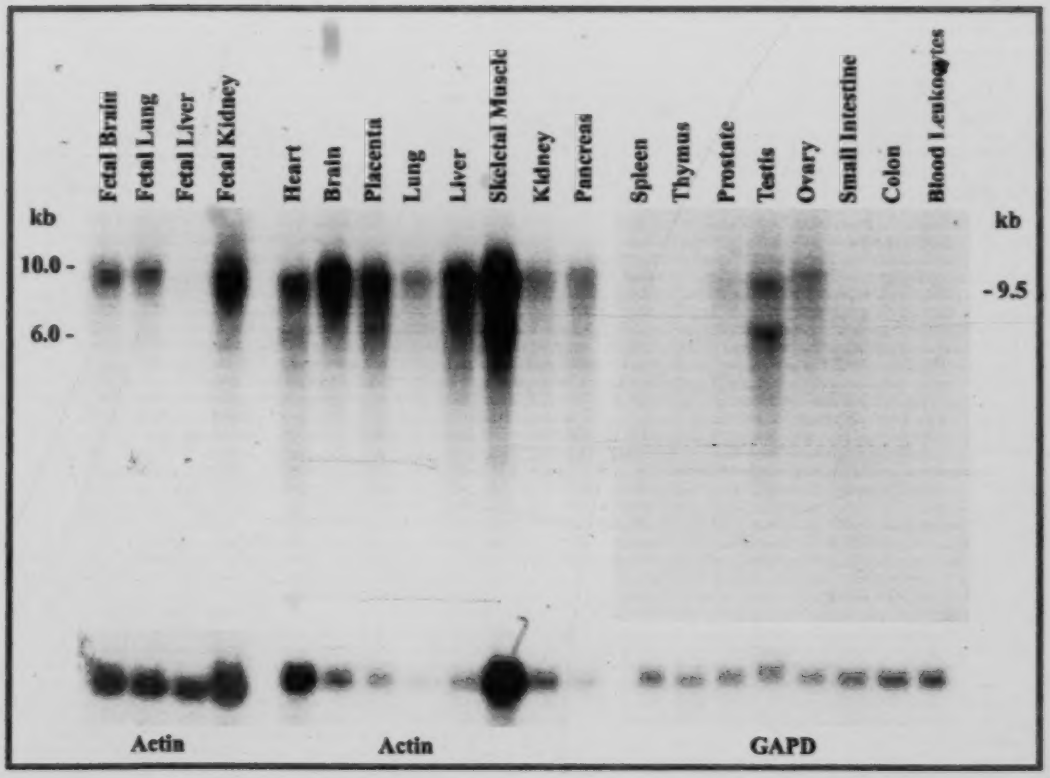


Figure 10. Expression profile of *CECR2*. Clontech™ human MTN northern blots were hybridized with a *CECR2* probe consisting of exons 1-10. Transcripts of 10.0, 9.5 and 6.0 kb are readily observable. Loading controls are shown below each blot.

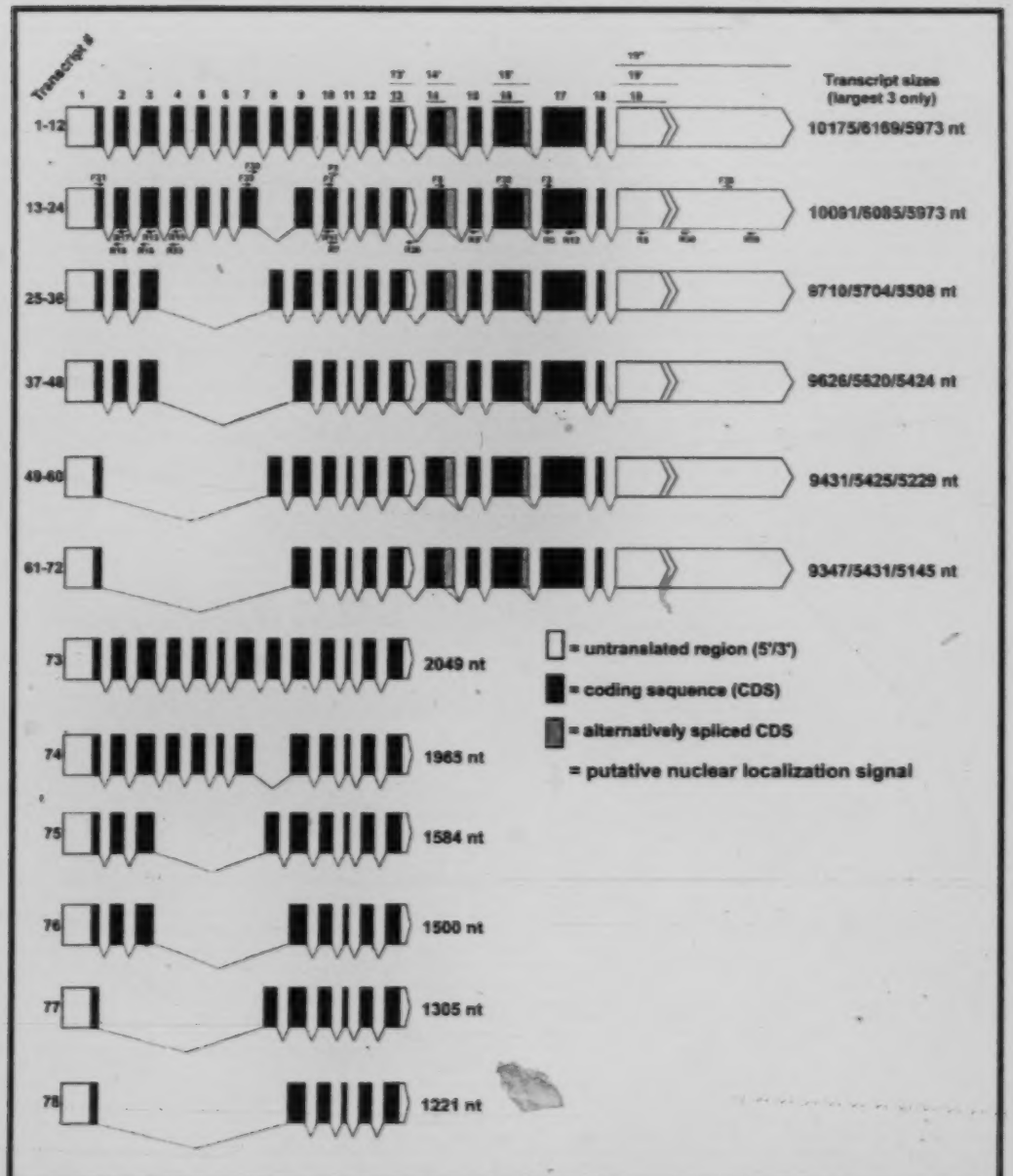


Figure 11. A subset of the theoretical transcripts of *CECR2*. Transcript sizes (nucleotides) are shown next to the transcript (or group of transcripts). For the first 6 groups (from the top) the different sizes correspond to which version of exon 19 is used. Not shown are sizes for the transcripts in which exon 14 (-84 bp), exon 16 (-165 bp) or exons 14 and 16 (-249 bp) are used instead of exons 14'and 16'. This adds a total of 54 more possible transcripts than what is shown for a total of 78. Note that in 52 of the 78 putative transcripts the putative nuclear localization signal would be spliced out.

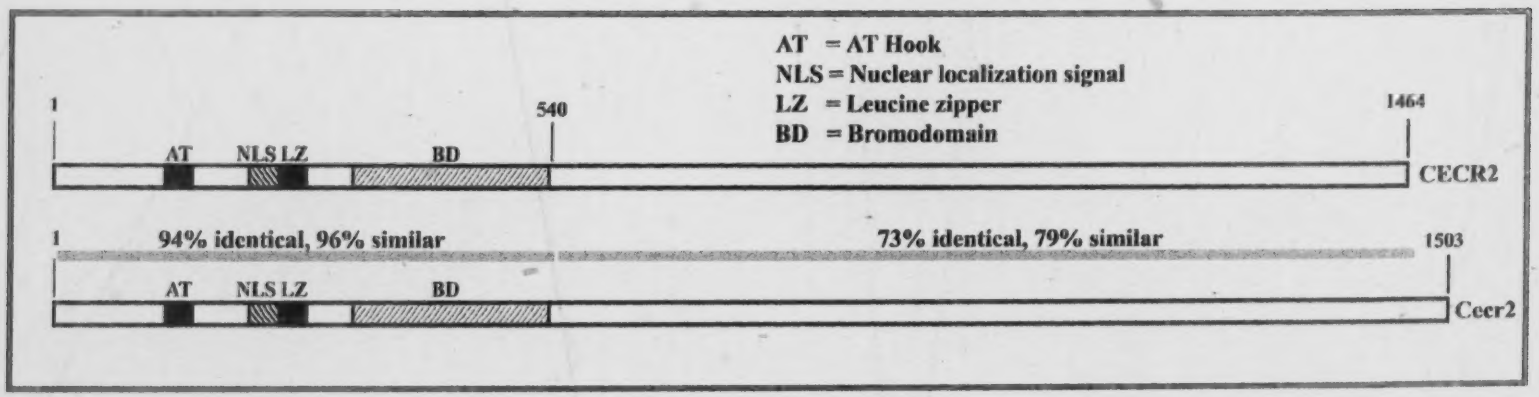


Figure 12. Comparison between the CECR2 and Cecr2 amino acid sequences. Note the high identity over the amino 1/3 of the proteins.

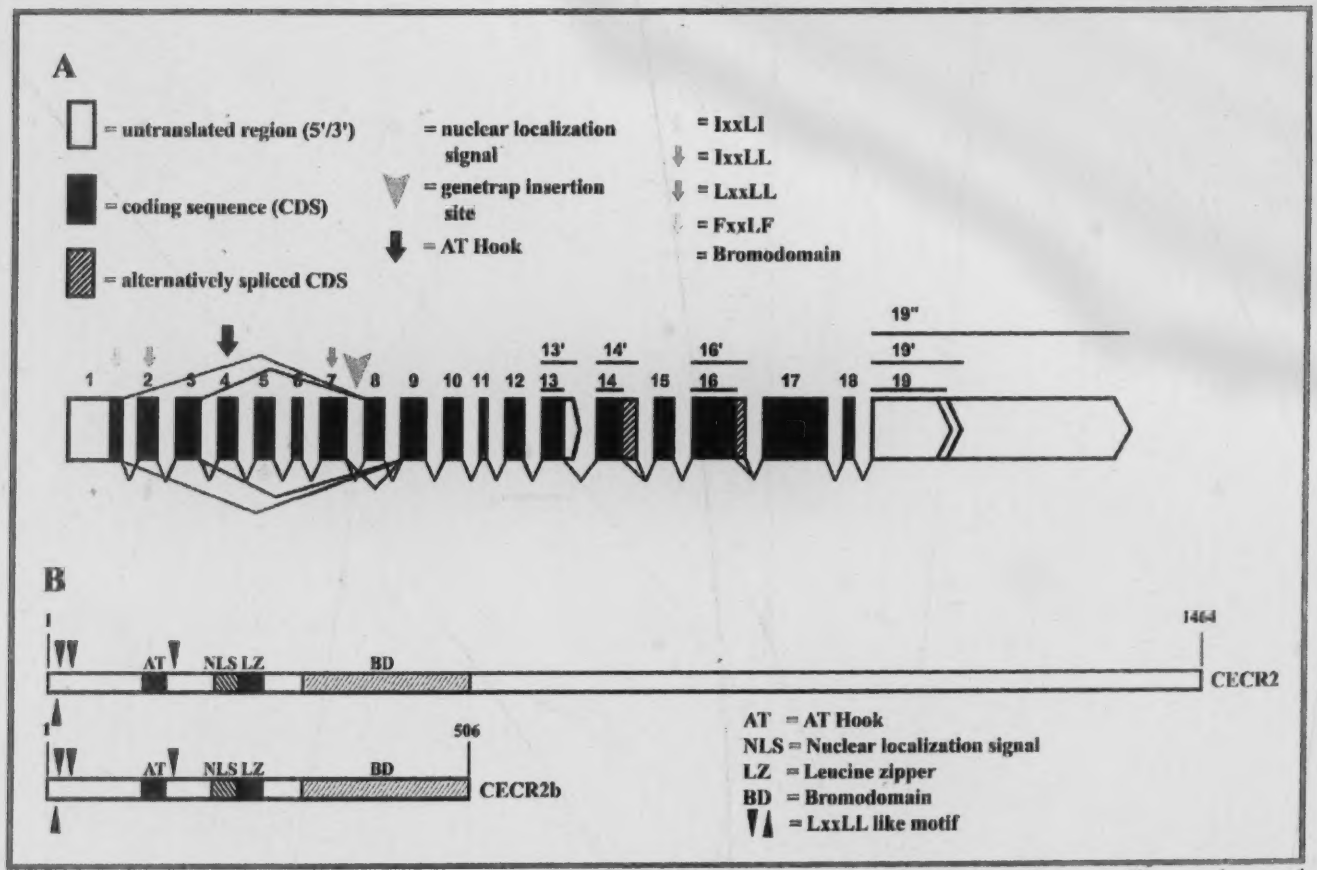


Figure 13. **(A)** Genomic structure of *CECR2* (not to scale). Location of motifs, untranslated regions, coding region and regions of alternative splicing are shown. The mouse locus (*Cecr2*) has the same structure with the exception of the absence of exon 13' **(B)** Hypothetical proteins derived from the *CECR2* locus. Only full length CECR2 and the shorter isoform CECR2b are shown, along with their associated motifs.

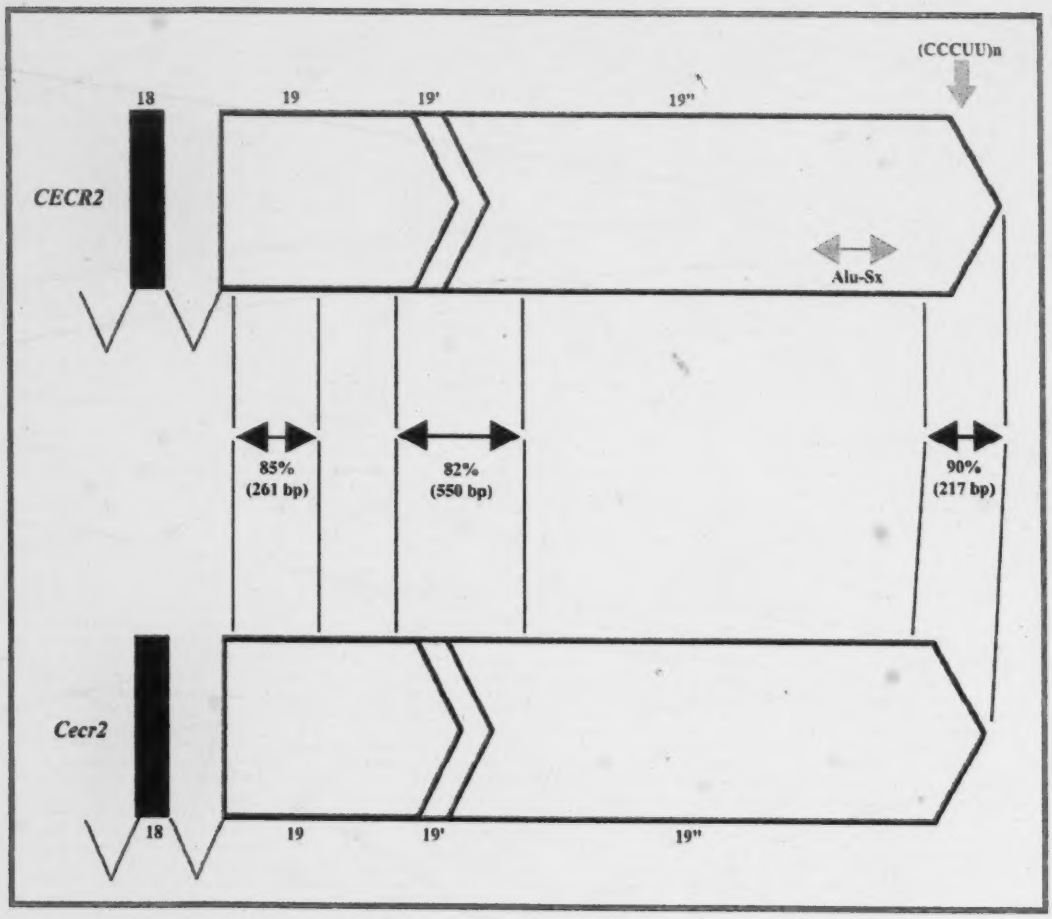


Figure 14. Sequence similarity between the 3' UTRs of *CECR2* and *Cecr2*. Regions 2 and 3 surround the three different polyadenylation locations (which produce exons 19, 19' and 19"). The locations of the Alu-Sx repeat and the $CCCUU_8$ repeat within *CECR2* are shown. Note that these two motifs are not present in *Cecr2*.

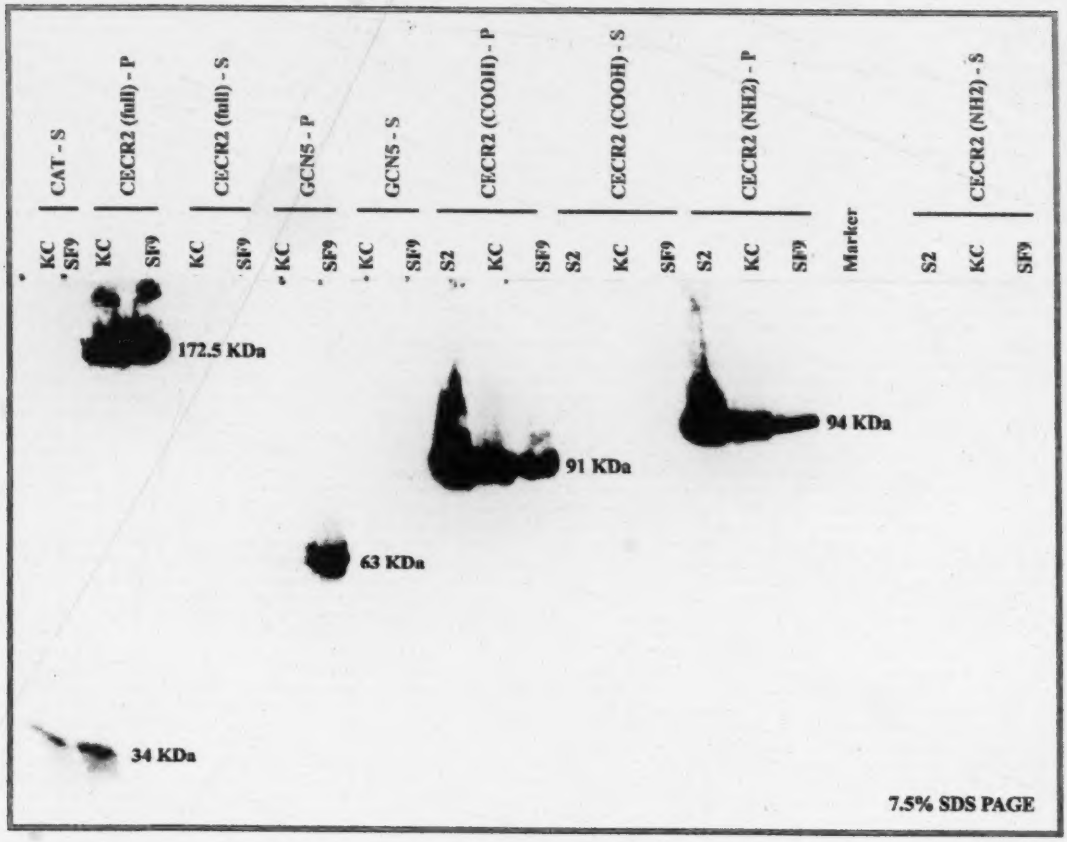


Figure 15. Western blot of human fusion proteins expressed transiently with an insect cell protein expression system. Cells were grown for 2 days post transfection and then harvested for protein extraction. KC and S2 are *Drosophila* cell lines and SF9 is a *Spodoptora frugiperida* cell line. P=cell pellet. S=media supernatant. CECR2-COOH = 2nd half of the CECR2 protein. CECR2-NH2 = 1st half of the CECR2 protein. CAT = chloramphenicol acetyltransferase (control protein). The blot was probed with an anti-V5 antibody.

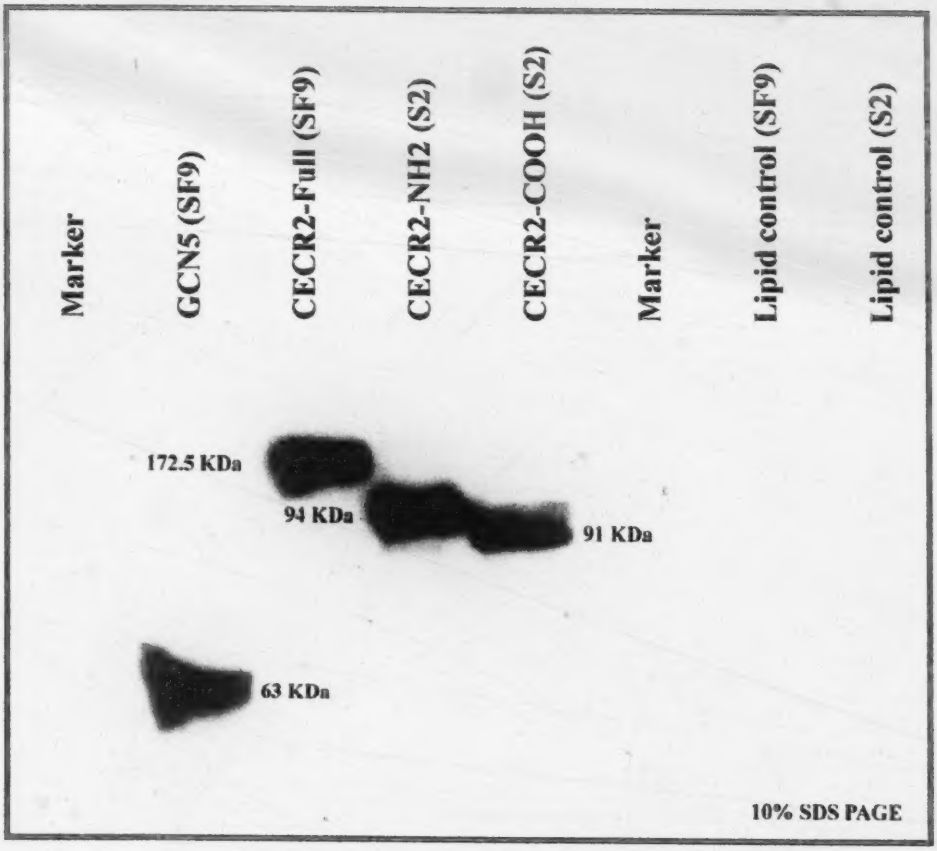


Figure 16. Western blot of human fusion proteins expressed with an insect cell protein expression system in stable cell lines generated by Blasticidin selection. All lanes correspond to protein extracted from the cell pellets. The blot was probed with an anti-V5 antibody. S2 is a *Drosophila* cell line and SF9 is a *Spodoptora frugiperida* cell line. CECR2-COOH = 2nd half of the CECR2 protein. CECR2-NH2 = 1st half of the CECR2 protein. Lipid controls correspond to cell lines transfected with liposomes only.

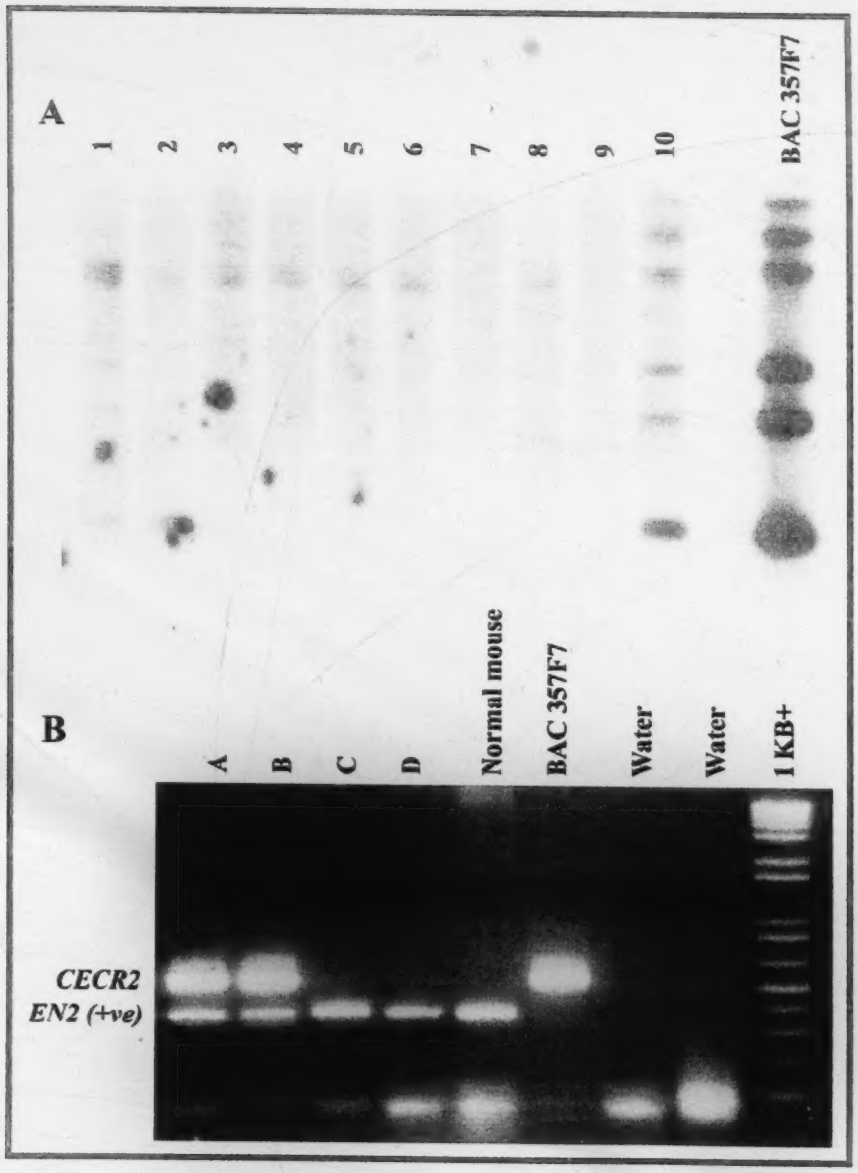


Figure 17. **(A)** Southern blot of *Eco*RI digested genomic DNA from the 10 initial mouse founders produced by pronuclear injection of BAC 357F7. Blot is hybridized with *CECR2* exons 1-10. Founders 1 and 10 show evidence of containing the transgene. Bands in founders 2-9 are likely due to cross hybridization with the native mouse locus **(B)** An example of the genotyping PCR that was used to track the BAC transgene. Four animals (A-D) and three controls are shown. *CECR2* specific primers (F26/R28, see **Figure 6** for location) amplify the transgene only. A mouse specific positive PCR control (from the *EN2* gene) is also included.

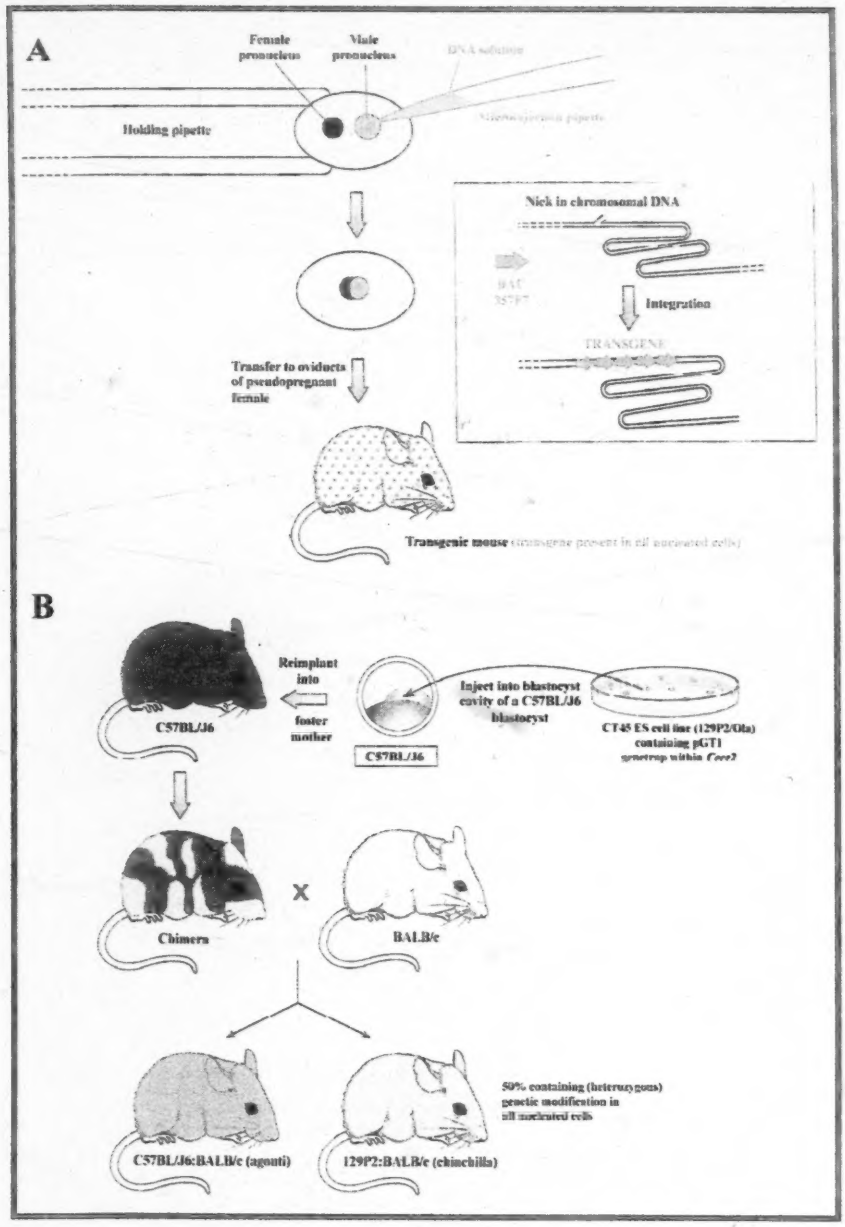


Figure 18. **(A)** Representation of transgenic creation via pronuclear injection of BAC DNA. **(B)** Diagram of the procedure used to generate a *Cecr2* knockout via ES cells containing a *Cecr2* genetrap. (Images modified from Strachan and Reid 1996)

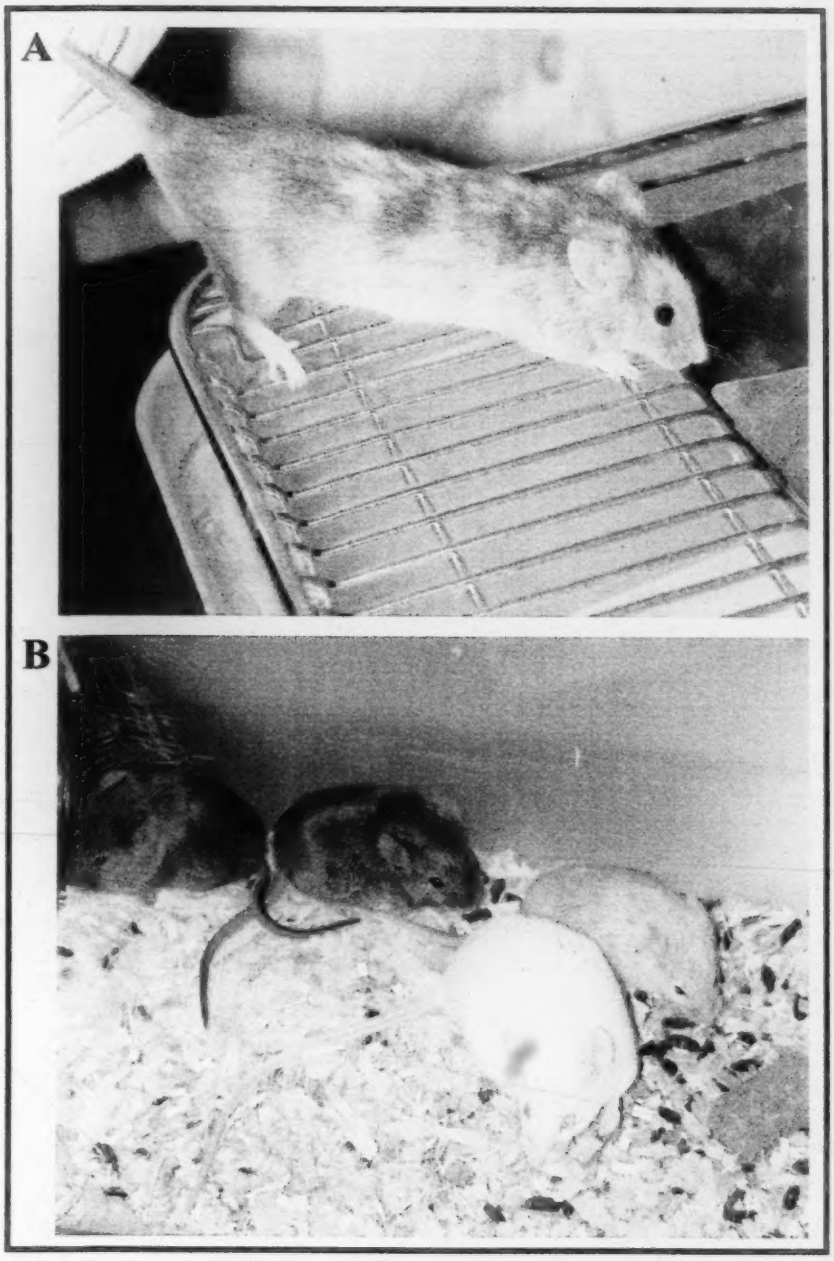


Figure 19. (A) The fertile male chimera produced from blastocyst injection of 129P2/Ola ES cells (containing a *Cecr2* gentrap) into C57BL/6J blastocysts. **(B)** F1 progeny of the chimera (in A) after mating to an albino BALB/c female. Sperm generated from C57BL cells of the chimera will produce agouti offspring when mated to a BALB/c female, while sperm from 129P2/Ola cells will produce chinchilla offspring (grey).

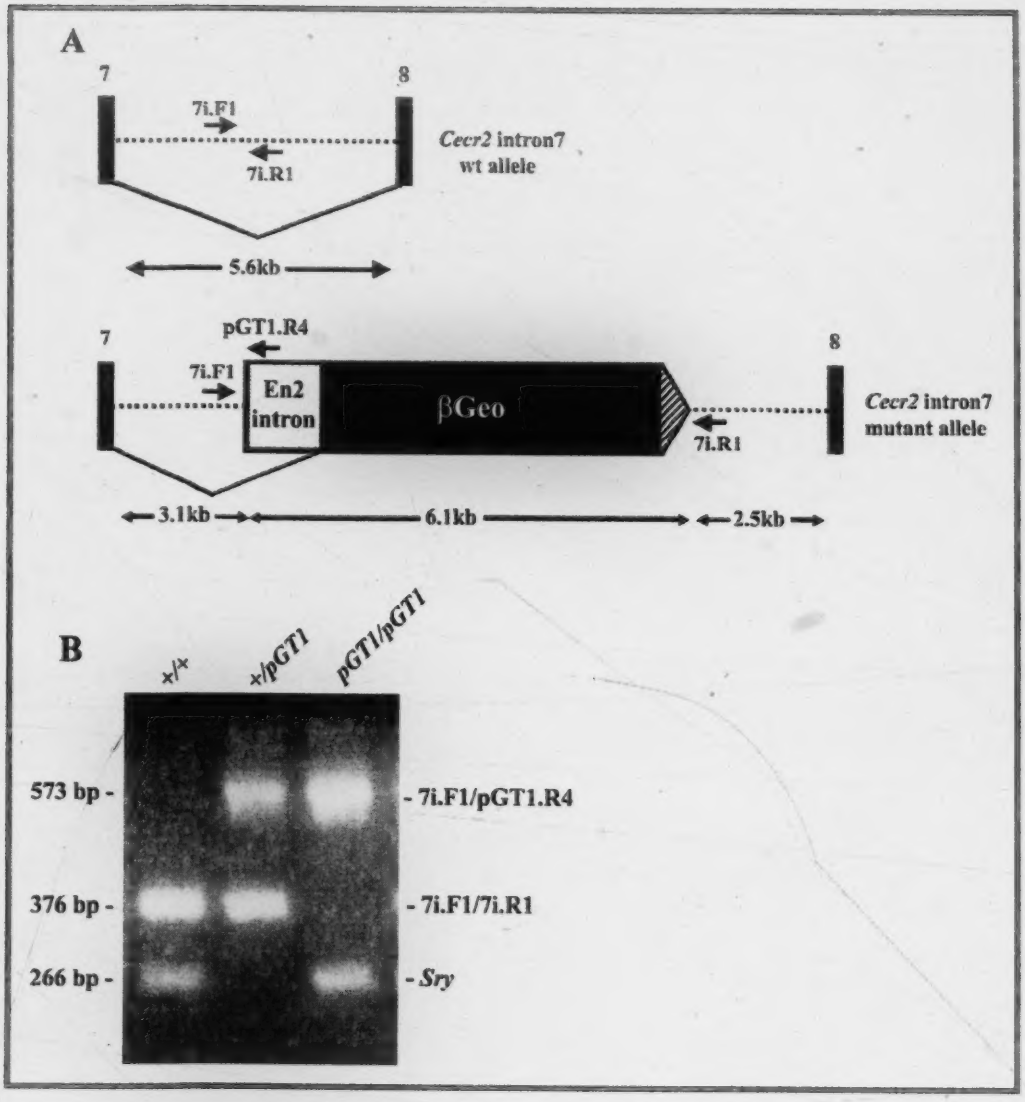


Figure 20. (A) Organization of the *Cecr2* intron 7 locus containing the pGT1 genetrap. **(B)** An example of the PCR genotyping used to track the locus.

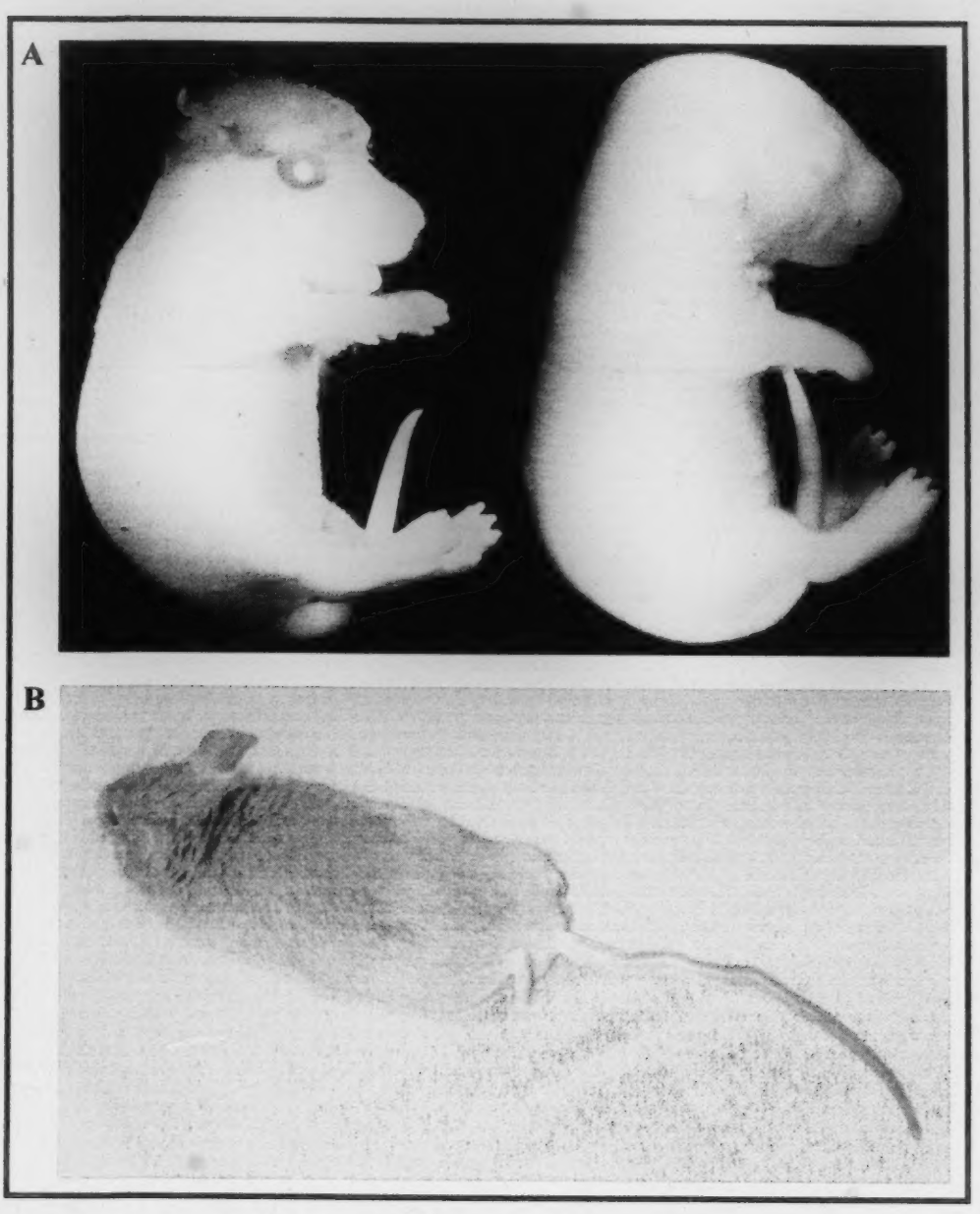


Figure 21. (A) Exencephaly phenotype that appears to segregate with the pGT1 genetrap. 19.5 dpc littermates: left = *Cecr2*$^{pGT1/pGT1}$, right = *Cecr2*$^{+/+}$. Penetrance is approximately 67% for homozygous animals. Also note the lack of an eyelid in the *Cecr2*$^{pGT1/pGT1}$ embryo. **(B)** Crooked tail phenotype of *Cecr2*$^{pGT1/pGT1}$ animals that are non-penetrant for exencephaly and are born. This phenotype is found at low penetrance (16% of homozygotes).

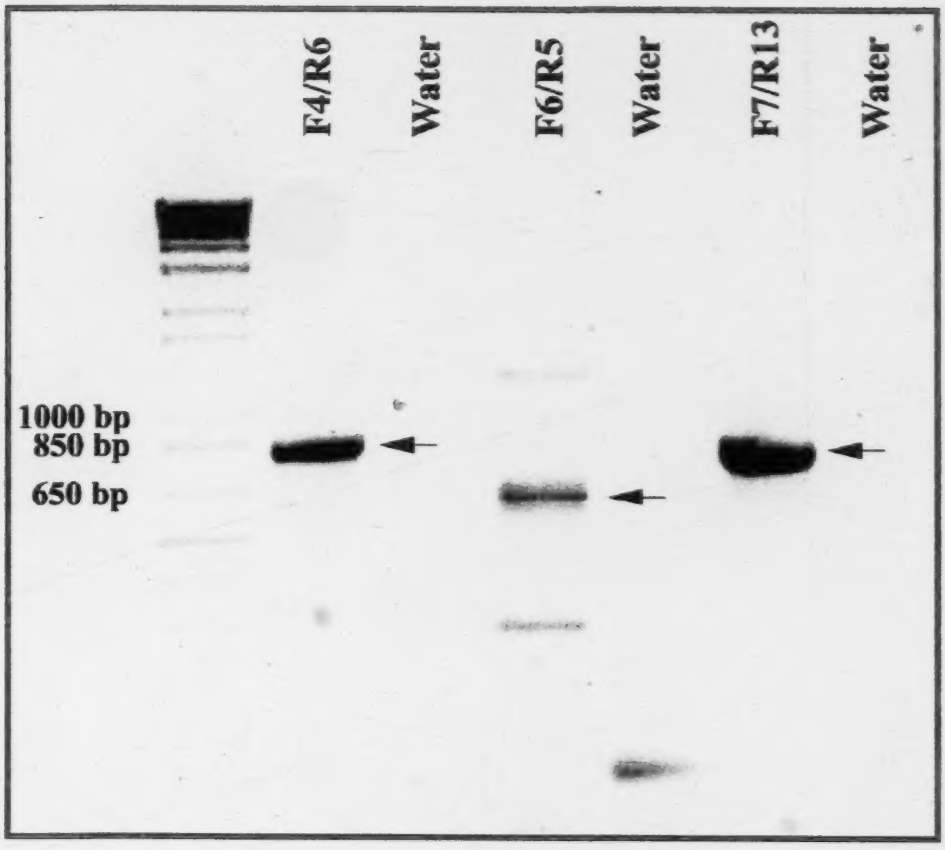


Figure 22. RT-PCR results of *Cecr2* on RNA collected from 14.5 dpc *Cecr2*$^{pGT1/pGT1}$ embryos. Three *Cecr2* primer sets (F4/R6 – exons 9-14'; F6/R5 – exons 14-16; F7/R13 – exons 16-17) 3' of the genetrap (intron 7) were employed to determine if readthrough was occurring. All the expected products were obtained, showing that the pGT1 genetrap allele of *Cecr2* is not a null.

Figure 23. Xgal staining of a 13.5dpc *Cecr2*$^{+/pGT1}$ embryo. Expression of the *LacZ* reporter gene is driven by the *Cecr2* promoter. Staining is apparent in the nasal cavity (grey arrow), forebrain (blue arrow), neural tube (purple arrow), spinal ganglia (black arrows), eyes (red arrow) and mesenchyme of the limbs (yellow arrow).

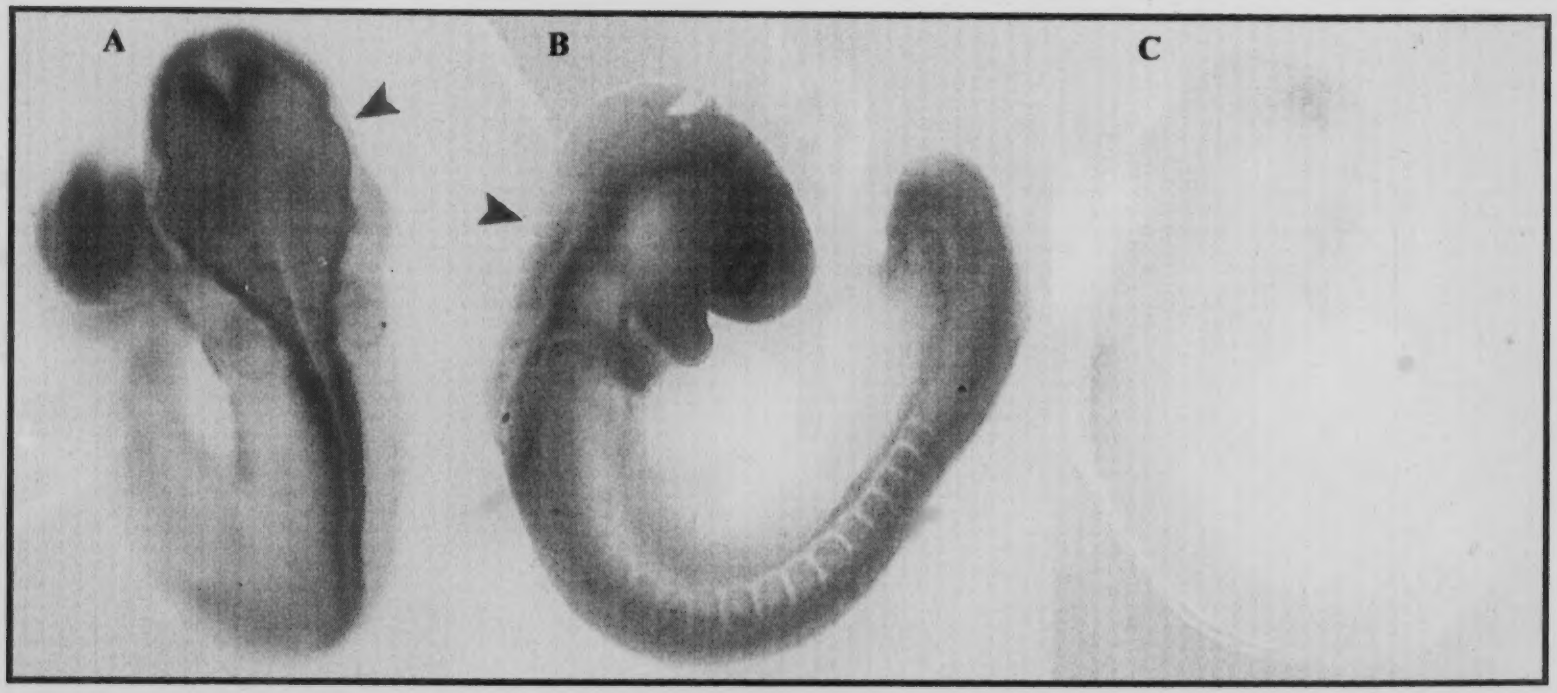


Figure 24. Xgal staining of a 9.5 dpc *Cecr2*$^{pGT1/pGT1}$ embryo (A, B) and a *Cecr2*$^{+/+}$ littermate (C). Expression of the LacZ reporter gene is driven by the *Cecr2* promoter. Note the open neural tube in the *Cecr2*$^{pGT1/pGT1}$ animal (purple arrow). Embryos provide by Tanya Ames.

Figure 25. Xgal staining of a 10.5dpc *Cecr2*$^{+/pGT1}$ embryo. Expression of the *LacZ* reporter gene is driven by the *Cecr2* promoter. Staining is apparent in the neural tube (purple arrow), spinal ganglia (black arrows), otic vesicles (red arrow), pharyngeal arches (orange arrows) and brain region (yellow arrow). Note the absence of expression in the heart region (blue arrow).

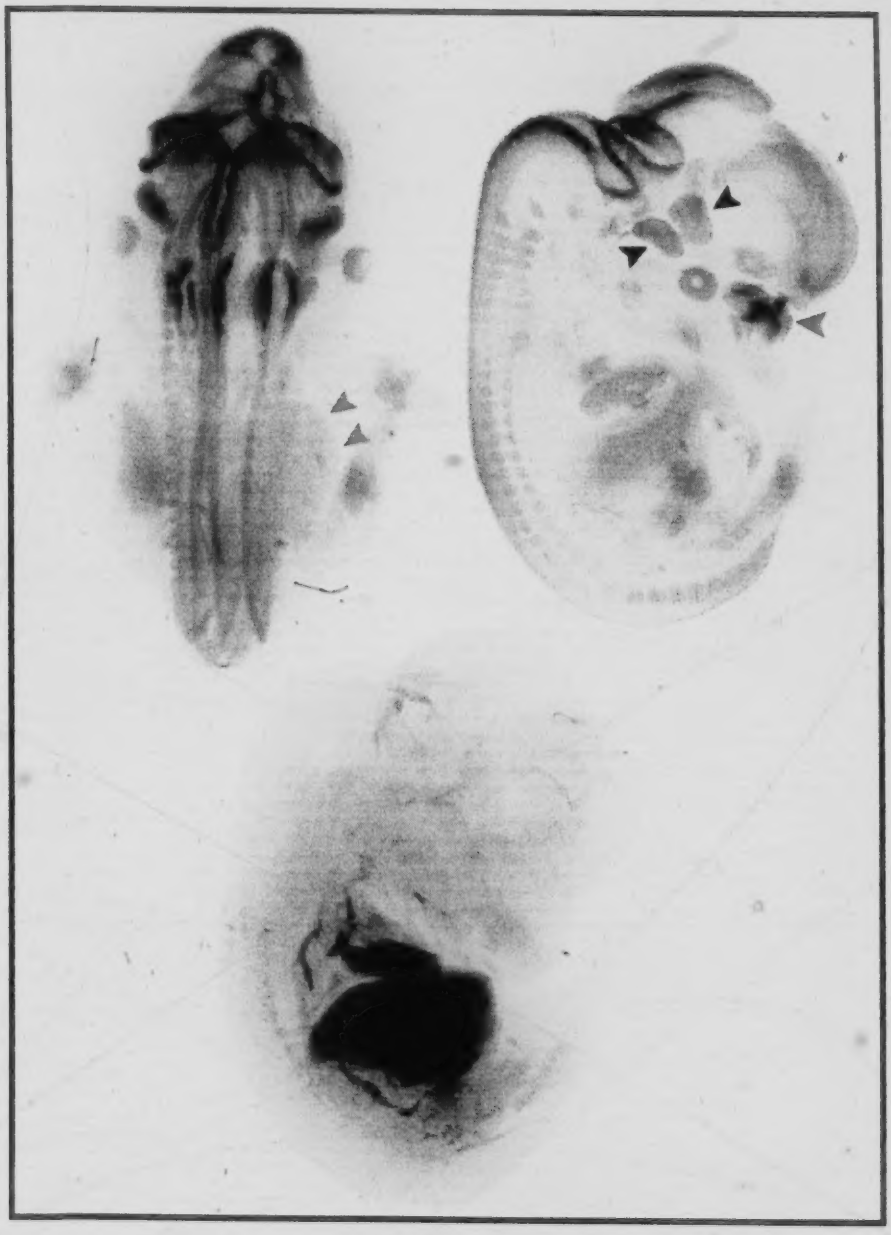


Figure 26. Xgal stained 13.5 dpc embryos. Top two embryos are *Cecr2*$^{pGT1/pGT1}$ that are non-penetrant for exencephaly, and the bottom embryo is a *Cecr2*$^{+/+}$ control. These embryos were stained in Xgal in the same fashion as those in Figures 23-25, but were subsequently cleared to better visualize interior structures. Note the staining in the intercostal region (red arrows), the large ganglia in the midbrain region (black arrows) and the nasal cavity (blue arrow) of the *Cecr2*$^{pGT1/pGT1}$ embryos.

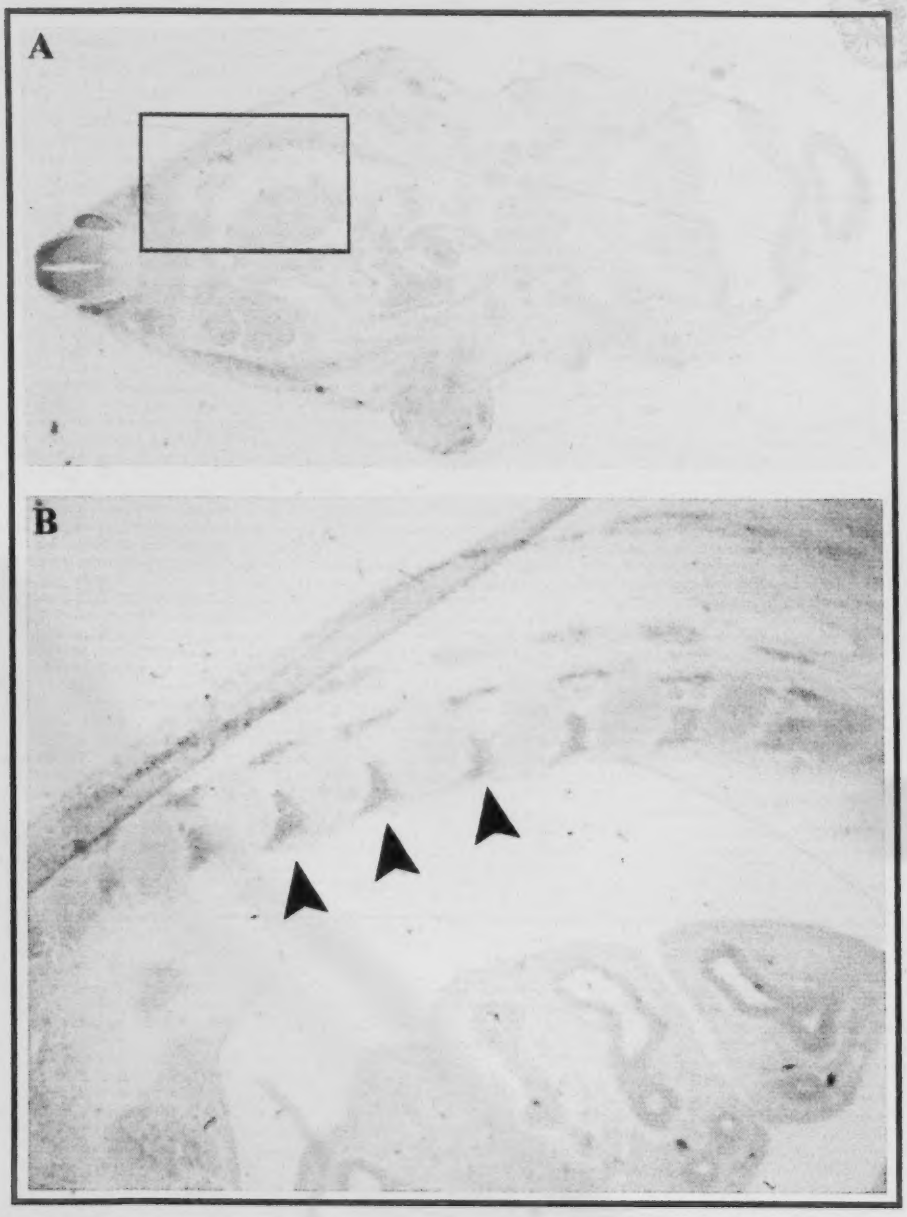


Figure 27. (A) Coronal section of 13.5 dpc *Cecr2*$^{+/pGT1}$ embryo stained with Xgal, showing *LacZ* reporter gene expression. **(B)** Enlargement of boxed region in (A). Note the β-*gal* staining in the intercostal regions (arrows). Sections provided by H. McDermid.

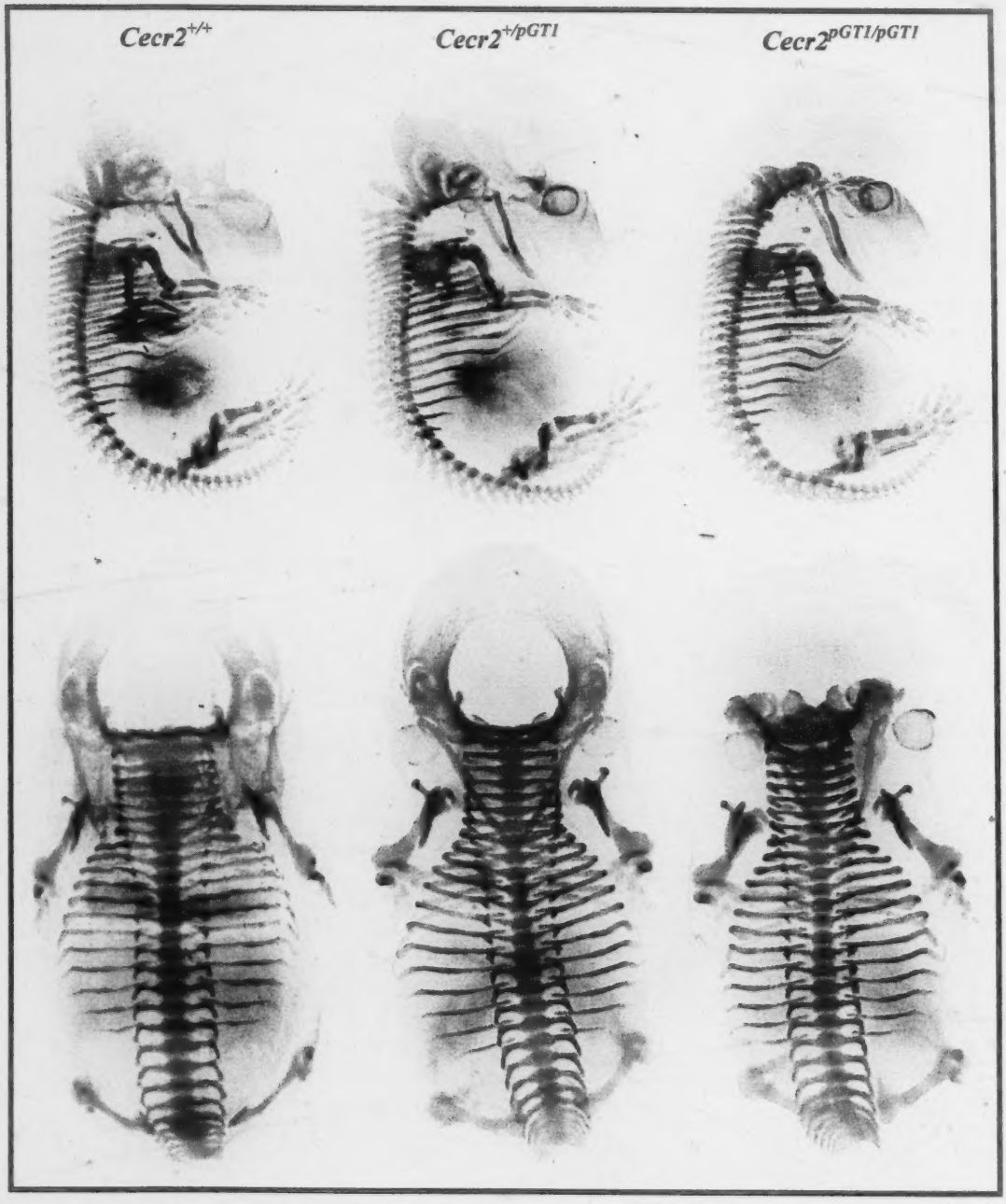


Figure 28. Skeletal analysis of *Cecr2* 14.5 dpc embryos stained with a cartilage specific dye. Lateral view (top) and dorsal view (bottom). The $Cecr2^{pGT1/pGT1}$ mice lack detectable cartilage development in the cranium, but are otherwise apparently normal.

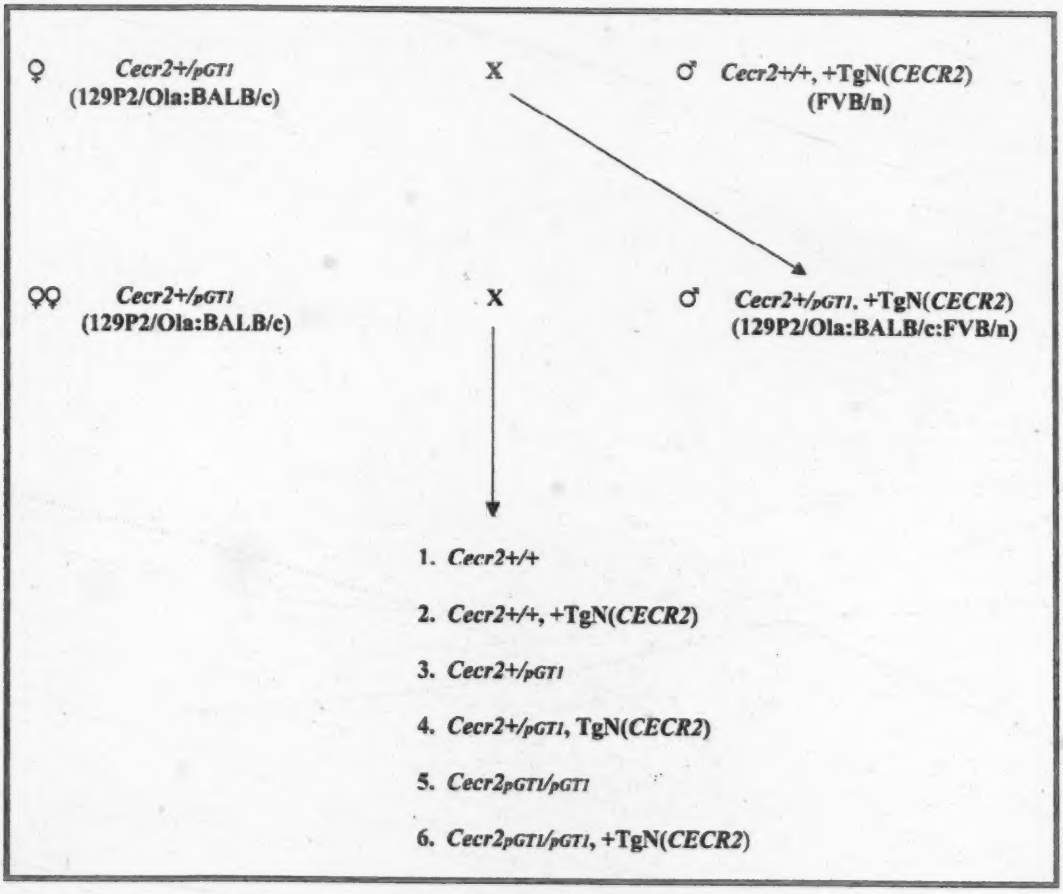


Figure 29. Diagram of the mating scheme employed in the *CECR2* transgene rescue experiment. A parallel control experiment was performed without the *CECR2* transgene (in mating number 2) in order to determine exencephaly penetrance on the 129P2/Ola:BALB/c:FVB/n mixed background.

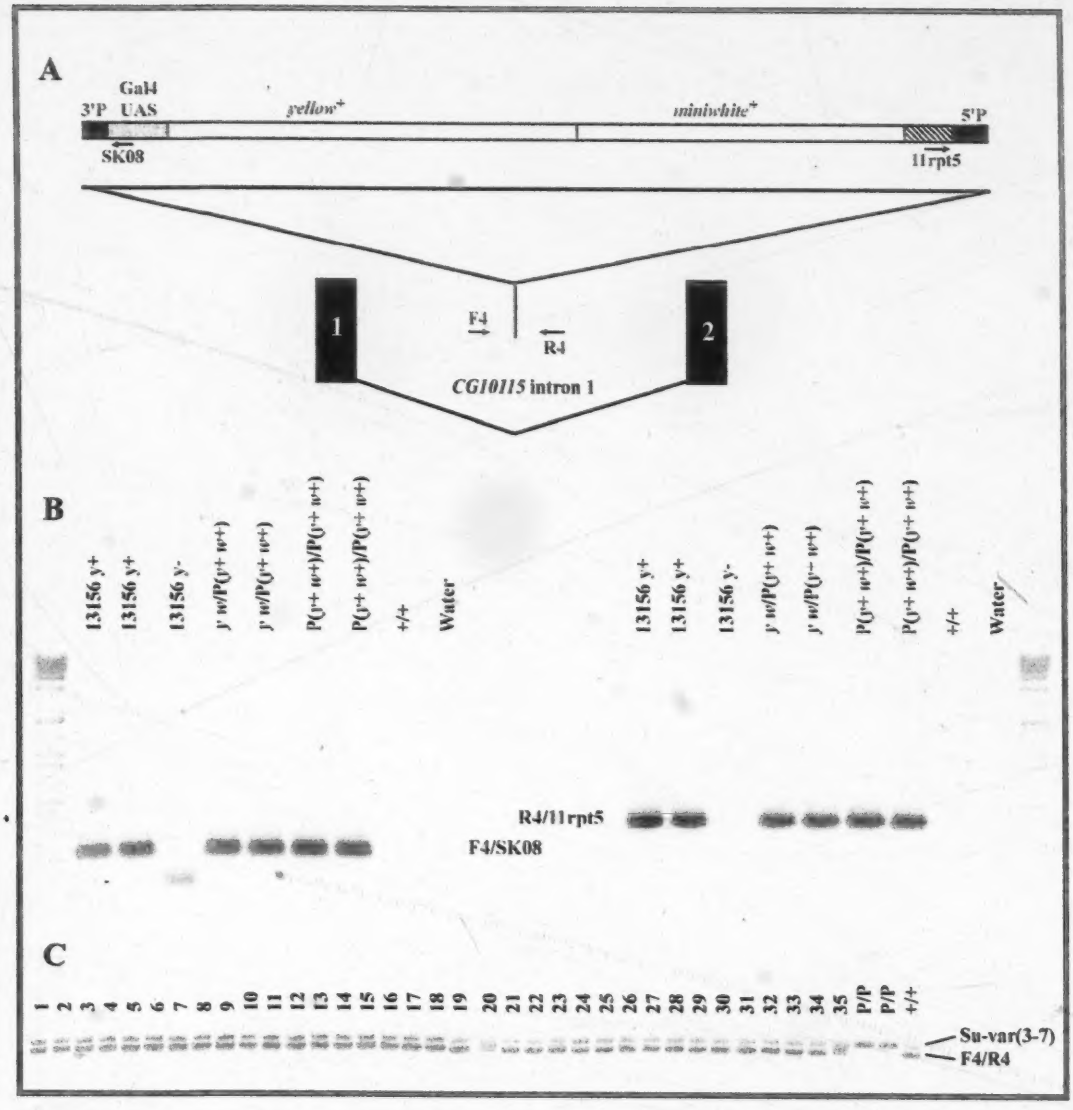


Figure 30. **(A)** Intron 1 of *Drosophila CG10115* in strain 13156, which contains a P-element insertion. **(B)** PCR between *CG10115* intron 1 and the P-element showing that it is in the reverse orientation with respect to the *CG10115* transcript. Neither wild type flies, nor *y*⁻ files found in the 13156 stock amplify with either primer set. (C) Single fly PCR (multiplex) of 35 P-element excision progeny. The positive control (Su(var)3-7) amplifies in all lanes to show that the PCR was successful. P-element homozygotes (P/P) fail to amplify F4/R4 as expected. An absence of the F4/R4 product in the excision progeny would indicate a deletion. No deletion mutants were observed.

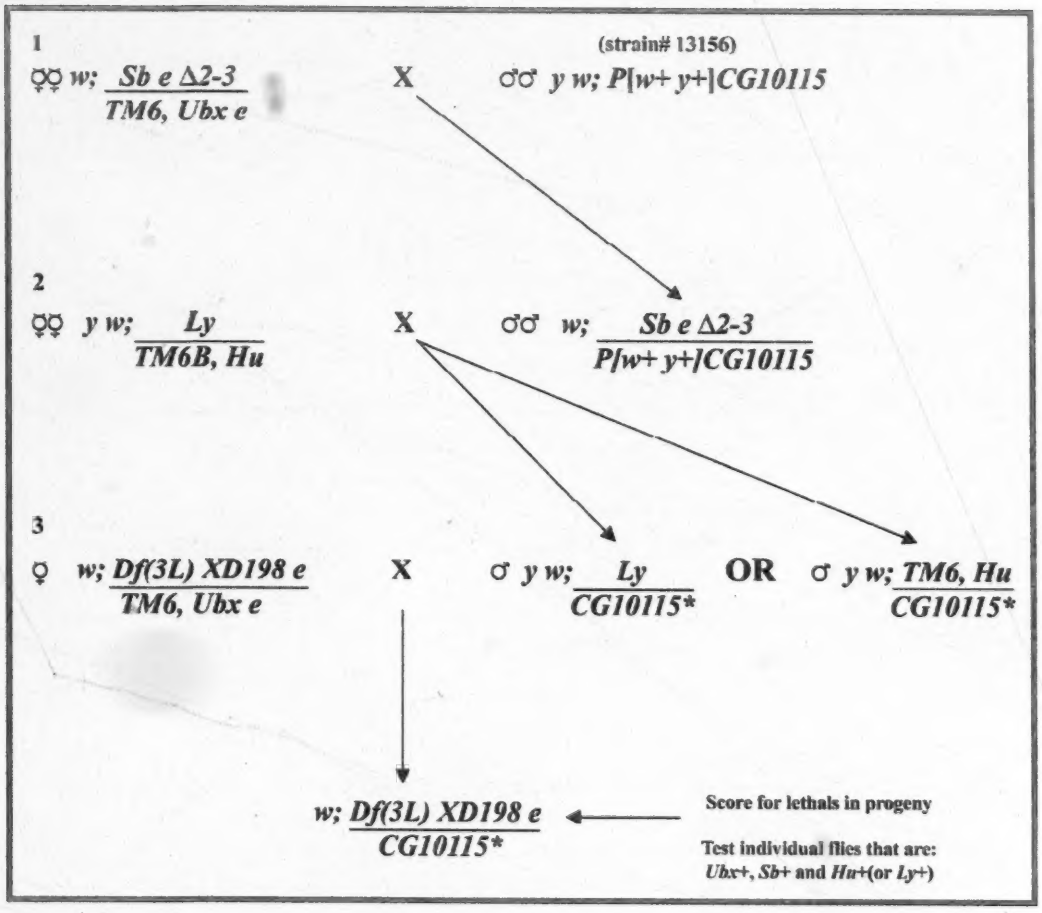

1
(strain# 13156)
♀♀ w; Sb e Δ2-3 / TM6, Ubx e
X
♂♂ y w; P[w+ y+]CG10115
2
♀♀ y w; Ly / TM6B, Hu
X
♂♂ w; Sb e Δ2-3 / P[w+ y+]CG10115
3
♀ w; Df(3L) XD198 e / TM6, Ubx e
X
♂ y w; Ly / CG10115*
OR
♂ y w; TM6, Hu / CG10115*
w; Df(3L) XD198 e / CG10115*
Score for lethals in progeny
Test individual flies that are:
Ubx+, Sb+ and Hu+(or Ly+)

Figure 31. Diagram of the *Drosophila melanogaster* crosses employed to mobilize the P-element located within the putative *CECR2* ortholog *CG10115*. The progeny of cross 3 were scored for lethals in the Ubx^{+}, Sb^{+} and Hu^{+} (or Ly^{+}) class. Progeny of this genotype (as no lethals were found) were tested by single fly PCR to determine if any deletions had been generated.

4. DISCUSSION

4.1 Identification of 3 genes in the distal CESCR

Cat eye syndrome results from a duplication, or triplication, of 22q11.2 (Schinzel et al. 1981; McDermid *et al.* 1986) and the identification of dosage sensitive genes in this region is the primary research focus of our lab. Through gene prediction, BLAST searching and comparative sequence analysis between human and mouse, I was able to identify three genes contained within two overlapping BAC clones at the distal end of the CESCR.

4.1.1 ATP6E, a subunit of vacuolar ATPase

ATP6E is one subunit of V1-ATPase, which is a multisubunit complex involved in the acidification of the vacuole (Baud *et al.* 1994). This gene had been previously cloned (Baud *et al.* 1994) but had not been finely mapped within the CESCR. Northern blotting analysis showed that this transcript is likely ubiquitously expressed and present as a single isoform. Mutations in other v-ATPase subunits typically lead to autosomal recessive disorders. For example, recessive mutations in *ATP6V0A4* (Smith *et al.* 2000) and *ATP6V1B1* (Karet et al. 1999) cause distal renal tubular acidosis. Based on mutational analyses of other genes with similar functions to that of *ATP6E*, it would seem unlikely that altered dosage of *ATP6E* would cause the severe developmental abnormalities found in CES, although it currently cannot be ruled out as a candidate gene. However, a transgenic mouse over-expressing *ATP6E* or *Atp6e* would address this question. Based on the rationale above, we chose not to further characterize *ATP6E* at this time.

4.1.2 SLC25A18, a mitochondrial glutamate transporter

SLC25A18 was cloned, and on the basis of similarity to other proteins, was categorized as a mitochondrial carrier protein. This large family of proteins shows homology between members and within internal regions of themselves, as these genes were generated based on an ancient gene amplification event(s) (Kramer and Palmieri 1992; Palmieri 1994). They reside in the inner or outer mitochondrial membrane and shuttle molecules across membranes. At the time of identification this gene, we were unable to determine the substrate that *SLC25A18* carried. However, it was subsequently determined by others (Fiermonte *et al.* 2002) that this protein GC2 (SLC25A18) is a H^+/glutamate antiporter found in the mitochondrial inner membrane. Of particular interest to us is that fact that a second homologue (GC1) was also identified and shown to have the same function. These two genes are 65% identical at the amino acid level, with weaker homology at the DNA level. Fiermonte *et al.* showed by real-time RT-PCR that *SLC25A18* (GC2) is expressed in multiple tissues, the strongest expression being in brain and testis, while other tissues have very low levels of transcript. Conversely, GC1 showed much higher expression in all tissues tested, with the strongest being pancreas, liver, brain and testis. I observed high expression in brain and liver for *SLC25A18* (GC2), but testis was not available on the Northern blots that I used. It is, therefore, possible that *SLC25A18* is indeed expressed in testis at a level detectable by Northern blotting, but this remains to be tested. My results for *SLC25A18* (GC2) expression in liver conflict with that of Fiermonte *et al.*, as I observed high expression in liver, whereas Fiermonte *et al.* showed minimal expression. The reason for this is unclear. A possibility for these

contradictory results in liver for GC2 could be explained by sample differences. If each experiment used liver tissue from one individual, possible differences in gene expression between two individuals should be considered. It is unlikely that the liver transcripts observed by Northern blot were the result of cross hybridization with GC1, as the level of DNA homology between the two is relatively low. In addition, if the liver signals were the result of cross hybridization, one would expect cross hybridization in other tissues where GC1 was highly expressed (e.g. pancreas), but this was not observed. One must take care in comparing data collected by two vastly different techniques (Northern vs. quantitative RT-PCR), as their sensitivity levels vary greatly.

Other members of the mitochondrial carrier superfamily are not typically dosage sensitive. For example, mutations in *SLC25A13* (Citrin) are recessive and lead to Citrullinemia (Ohura *et al.* 2001), while homozygous mutations in the mitochondrial deoxynucleotide carrier *SLC25A19* lead to Amish lethal microcephaly (Rosenberg *et al.* 2002). The limited expression profile of *SLC25A18* would also argue against involvement in CES, as it is not expressed in the majority of CES-affected tissues. Secondly, considering there are already effectively 4 copies in the genome (2 GC1 and 2 GC2), *SLC25A18* (GC2) is an unlikely candidate for dosage sensitivity. Furthermore, *SLC25A18* is contained in its entirety within BAC357F7, suggesting either that it is not dosage sensitive or that this gene is not functional in mouse, as mice carrying this transgene appear normal. However, a more detailed expression analysis of *SLC25A18* in the BAC357F7 transgenic mice is warranted to determine if it is expressed appropriately. Based on the evidence above, we do not feel that *SLC25A18* is a good candidate for CES involvement at this time.

4.1.3 CECR2, a putative transcriptional regulator

CECR2 was cloned in multiple segments due to its large size. It spans approximately 200 kbp of BAC 357F7, approximately 107 kbp of which is comprised of intron 1. The predicted protein is 1464 amino acids, with an unmodified mass of 173 kDa. There are multiple regions of significant identity between human and mouse within intron 1 (Footz *et al.* 2001), but it is not clear as to their function. RT-PCR was attempted between several conserved regions and exon 2 to determine if these were alternative first exons, but products were not obtained in any of the tissues tested. Thus, it is possible that they are tissue or temporally specific alternative exons, or that they are regulatory elements required for the expression of *CECR2*. It was also determined that the Genbank sequence file for BAC 357F7 was missing a region surrounding the first exon of *CECR2*. This region was cloned and sequenced to confirm *CECR2* exon 1. Other members of the lab have also identified sequence "deletions" in Genbank sequence files. In consideration of these findings, one must take care when utilizing Genbank sequence files of large genomic clones, which have been "shotgun" sequenced. It is likely that the vast majority of sequences are correct, but one must be prepared for the possibility of errors.

4.2 *CECR2* encodes a protein likely involved in transcriptional regulation and/or chromatin remodeling

Sequence analysis of *CECR2* revealed several interesting motifs that would be expected to be present in a nuclear protein that associates with chromatin and DNA. The most informative amino acid motif in CECR2 is the bromodomain, which has been

previously shown to interact with acetylated lysine residues on the amino terminal tails of histones (Dhalluin *et al.* 1999). A putative nuclear localization signal in exon 5 of *CECR2* further corroborates this possibility. In addition, the AT hook motif, found in exon 4, has been previously shown to physically interact with AT rich regions of the minor groove of DNA (Reeves and Nissen 1990; Aravind and Landsman 1998), suggesting that CECR2 may directly interact with DNA, in addition to histones. These 3 motifs are all suggestive of a role for CECR2 in the nucleus on DNA, acting as a transcriptional regulator. Cecr2-DNA contact is again supported by results of a *Cecr2* genetrap ES cell line (Tate *et al.* 1998). Tate *et al.*, through a gene trapping approach, tagged *Cecr2* with a LacZ reporter gene. The pGT1 genetrap splices in-frame to Cecr2 after exon 7. This fusion would leave the NLS and AT hook intact, but exclude the bromodomain motif. The observation of the Cecr2-βgal fusion protein associated with DNA and/or chromatin does support the presence of a DNA or chromatin binding motif within the first 7 exons of Cecr2.

Immunostaining with β-galactosidase antibodies showed that the tagged form of Cecr2 is associated with heterochromatin during metaphase in ES cells, and was localized to the outer sides of each chromatid (Tate *et. al.* 1998). During other portions of the cell cycle, the Cecr2/β-gal fusion protein had diffuse nuclear staining in addition to concentrations along the length of the chromosomes, excluding the centric heterochromatin (Tate *et al.* 1998). Tate *et al.* also report Cecr2 to be localized to heterochromatic regions during interphase, in addition to diffuse nuclear staining. It would appear that Cecr2 subnuclear localization is dynamic and varies according to the cell cycle. This may be suggestive of distinct roles for Cecr2 based on cell cycle timing. However, the absence of the bromodomain in the fusion protein may dramatically alter

the localization of Cecr2. The ES cell immunohistochemistry results of Tate *et al.* should be viewed with scrutiny, as this may not reflect the true localization pattern of Cecr2 due to the truncation of the protein. Generation of Cecr2 specific antibodies and subsequent immunohistochemistry in wildtype cells should allow a definitive subnuclear localization to be determined.

The possibility of CECR2/Cecr2 having histone acetyltransferase (HAT) activity is plausible considering that several other proteins (CBP, p300, GCN5, P/CAF) containing bromodomains also exhibit HAT activity (see *1.4.2* for references). However, it should be noted that not all bromodomain proteins are HATs. I was unable to determine if CECR2 displays HAT activity due to time and technical constraints. Testing CECR2 for histone acetyltransferase, or deacetylase activity (see below) is, however, certainly an avenue worth examining in the future. All information to date regarding *CECR2/Cecr2* is suggestive of a role as a transcriptional regulator.

Recently, it has been shown that CECR2 forms part of an ATP dependant chromatin-remodeling complex, which includes CECR2, SNF2L, and HBXAP (Ramin Shiekhattar, personal communication). HBXAP is a PHD (plant homeodomain) finger containing protein that has been shown to possess transcription repression activity (Shamay *et al.* 2002). This may be suggestive of this CECR2 containing complex being repressive in nature. Previous evidence has shown that the bromodomain and PHD finger of the transcription co-repressor KAP-1 function cooperatively in transcription repression via recruitment of the NURD subunit Mi-2α (Schultz et al. 2001). NURD is a multisubunit protein complex that possesses both ATP-dependant chromatin-remodeling activity and histone deacetylase activity (Xue *et. al.* 1998). Thus, it is possible that the

bromodomain of CECR2 and the PHD finger of HBXAP act cooperatively in a similar manner to KAP-1. This would also suggest the possibility of histone deacetylase activity in CECR2 or the other members of the CECR2 complex. Alternatively, the CECR2/HBXAP/SNF2L complex may further recruit additional proteins to modify histones. Further investigation of potential covalent histone modification activities of CECR2, HBXAP and SNF2L are, thus, warranted.

The other member of the CECR2 complex, SNF2L, is a relatively uncharacterized protein. It is a homolog of the *Drosophila Imitation Switch* (*ISWI*) gene and the yeast ISW1/ISW2 proteins, along with its paralog, SNF2H (Okabe *et al.* 1992; Aihara *et al.* 1998; Tsukiyama *et al.* 1999). *Snf2l* expression has been shown throughout the embryo during embryonic development (Lazzaro and Picketts, 2001). Several different mRNA transcripts are seen, including a brain specific transcript, which is up-regulated after birth (Lazzaro and Picketts, 2001). *Snf2l* expression is comparatively lower than *Snf2h* during development, but the reciprocal is true postnatally (Lazzaro and Picketts, 2001). Presently, it is not clear as to the number of complexes in which SNF2L/Snf2l resides, and whether SNF2L/Snf2l levels are critical during development. One could surmise, however, that Snf2l levels may be crucial based on the Cecr2 results presented here, which potentially implicate Cecr2 in neurulation. Therefore, based on the observation that SNF2L and CECR2 are present in a common complex, the investigation of a potential role for SNF2L/Snf2l in neurulation seems prudent.

Since the AT hook, nuclear localization signal, and bromodomain are all found within the first third of the protein (and also in the smaller *CECR2b* transcript), the potential function of the remaining two thirds of the CECR2 protein remains unclear. The

region between exons 14 and 19 comprises 965 amino acids in human and 953 amino acids in mouse. Over this region the proteins are 74% identical and 79% similar between human and mouse, as compared to 94% identical and 97% similar over the first 13 exons (498 amino acids in both human and mouse). Clearly, there are some selective constraints upon the first 13 exons of *CECR2/Cecr2* that are not as strong in the remaining 2/3 of the protein. I believe this is due to protein function that is associated with these regions. There is likely selective pressure on the first third of the protein such that it will not lose the ability to bind DNA or histones, which are highly conserved over evolution. As for the remainder of the protein, there are no informative motifs, with the exception of a putative ATP binding pocket. With respect to the lower homology in the carboxy 2/3 of *CECR2/Cecr2*, it is possible that the region is adapting to changes in other molecules in order to maintain interactions. Transcriptional regulators are notorious for interacting with many different proteins, and the possibility that these partner proteins are evolving could potentially put pressure on regions of CECR2/Cecr2 to evolve in a coordinated manner. Potential interaction domains with SNF2L and HBXAP have not been shown, but the carboxy region of CECR2 may include such regions. The carboxy 2/3 of CECR2/Cecr2 is clearly important in some manner as it is conserved between human and mouse at a reasonable level (~80%), yet there is also a transcript (*CECR2b*) that omits this region entirely. The rationale for expressing this truncated transcript remains unknown at this time. In addition, it is not clear as to whether the *Drosophila* homologue (*CG10115*) contains the distal region (equivalent of exons 14-19) in its transcript. If it does, it is not conserved between the human or mouse transcripts. However, there clearly is conservation within the first 500 amino acids of the human/mouse proteins to the

putative *Drosophila* gene. The *Drosophila CG10115* locus will require more study in order to address these questions.

4.3 *CECR2* has a complex expression pattern

As addressed in the previous section, the *CECR2* locus is complicated. There is a short transcript composed of exons 1-13' that produces a protein that is effectively 1/3 of the size of the full-length protein encoded by exons 1-19. In addition, there is alternative splicing within portions of exons 14 and 16, and splicing that removes exons 2 through 8 in various combinations. The reason for these splicing events is not clear. Each alternative splice variant was found in all tissues that were tested (relative amounts were not determined), but the possibility exists for tissue specificity at different developmental time points that were not tested. RNase protection assays may allow a better determination of the true levels of the various *CECR2* alternative transcripts, and these experiments are currently under development. Using current techniques we were unable to determine any correlation between different alternative splice events and different alternative polyadenylation events. Specifically, are certain splice variants only/preferentially found in combination with certain polyadenylation events? The alternative splicing within the *CECR2* locus is conserved in mouse; this conservation suggests a biologically significant role(s). It is plausible that some of these regions of splicing, such as the 3' ends of exons 14 (28 aa) and 16 (55 aa), could contain regions that interact with primary partners. It is also possible that they also disrupt a second protein interaction domain. Therefore, when these regions of exons 14 and 16 are removed, a secondary protein interaction domain could be reconstituted, allowing

CECR2 interaction with novel partners. These are clearly areas of the protein that would be interesting to study with respect to potential protein-protein interactions.

Addressing the question of the function of a CECR2 protein missing the regions encoded by exons 2-8, 4-8 or 8 alone has been intriguing. Any splice variant that removes exons 4 and 5 would remove the putative AT hook and nuclear localization signal, generating a potentially cytosolic protein. The possibility of a secondary nuclear localization signal cannot be ruled out, yet these variants are clearly significant in the removal of a highly conserved region of the protein in human and mouse. As for the potential function of a cytosolic version of CECR2, one possibility is that if CECR2 does in fact have acetyltransferase activity, it may activate proteins other than histones by acetylation in a manner similar to phosphorylation. Several transcription factors and coactivators are acetylated, regulating their function. ACTR (Chen *et al.* 1999b), p53 (Barlev *et al.* 1999), and SREPB1a/2 (Giandomenico *et al.* 2003) are all acetylated by CBP, whereas steroidogenic factor 1 is acetylated by GCN5 (Jacob *et al.* 2001). Subcellular redistribution, between the nucleus and cytoplasm, of a coactivator has also been shown with p/CIP (Qutob *et al.* 2002). Thus, the potential for a cytosolic form of CECR2 should not be ignored. Immunohistochemistry and subcellular localization studies are clearly required to determine if CECR2 does have cytosolic forms.

In addition to alternative splicing, there are several alternative polyadenylation events within *CECR2* and *Cecr2*. *CECR2b* has a unique extended version of exon 13 (exon 13'), which adds 7 new amino acids and changes one amino acid bounding the splice site to exon 14, followed by a short (~50 bp) 3' UTR. Exons 14 through 19 are not a part of this small transcript (*CECR2b,* 506 aa). This short transcript generates a protein

approximately 1/3 the size of the full length CECR2 yet retains all the recognizable protein motifs (AT hook, NLS, bromodomain, LxxLL motifs). Exon 19 has 3 different locations for polyadenylation, which are all confirmed by ESTs, with two being relatively close together, producing exon 19 sizes of 1.1 and 1.3 kb. Of particular interest is the polyadenylation event that occurs much further 3', producing exon 19" that is 5.3 kb. As the stop codon is bases 4-6 of this exon, it only contributes 1 amino acid to the final CECR2 protein and, as such, it is almost entirely untranslated. The presence, but not sequence, of the large 5.3 kb version of exon 19 is conserved in mouse, suggestive of a role in the regulation of the mRNA. As signals in mRNA are difficult to determine based on primary sequence alone, I was unable to determine which, if any, of the regions of this exon have regulatory effects on the mRNA. In addition, Northern blots were particularly "smeared" in our results and the results of others (Liu and McKeehan 2002), with only the largest transcripts appearing consistently. This suggests that the *CECR2* mRNA may be unstable. The presence of the $CCCUU_8$ repeat in *CECR2* is intriguing, as similar CT rich motifs have been shown to bind to HuR (Yeap *et al.* 2002), a protein which shuttles between the nucleus and cytoplasm, stabilizing mRNAs by physically interacting with them (Levy *et al.* 1998; Fan and Steitz 1998a; Fan and Steitz 1998b). This may explain why transcripts with exon 19" seem to be the most prominent on Northern blots. It remains to be tested, however, if this $CCCUU_8$ motif has any effect on *CECR2* RNA stability. The fact that mouse *Cecr2* does not contain the $CCCUU_8$ is interesting insofar as *Cecr2* is not particularly amenable to *in situ* hybridization in mouse embryos, or to Northern hybridization. If this motif is required for RNA stability, and is not present in mouse, this could suggest a potentially short half-life for *Cecr2* in mouse. If there is a

short half-life for *Cecr2* mRNA, then RNA characterization will be difficult, and it would perhaps be more feasible to characterize the Cecr2 protein by immunohistochemistry. This will, however, require use of a Cecr2 antibody, preferably one that does not cross-react with CECR2. Cecr2 antibodies are currently under development.

The rationale for potential differential regulation of *CECR2* vs. *Cecr2* is not clear at this time. The $CCCUU_8$ motif is found in several ESTs in the human genome, in addition to rat and mouse, so the potential for a biological function of this motif cannot be overlooked. The other intriguing possibility with the $CCCUU_8$ motif is the potential involvement with nuclear/cytoplasmic shuttling via interaction with the HuR protein, which is known to shuttle between the nucleus and cytoplasm (Fan and Steitz 1998a; Fan and Steitz 1998b; Yeap et. al. 2002). Versions of *CECR2* mRNA lacking the NLS could potentially move into the nucleus, via association with HuR. It would certainly be interesting to use the three different versions *CECR2* 3' UTR in an *in vivo* experiment in which it is attached to a reporter gene coding region in order to determine if the reporter gene's mRNA is, indeed, affected in any way, such as turnover or localization. Protein binding studies of the *CECR2* 3' UTR would also be of interest, specifically to confirm whether or not HuR binds to the 3' UTR of *CECR2*.

4.4 *CECR2* over-expression in mouse produces conflicting results

We generated (FVB/n) transgenic mice carrying the human BAC 357F7 transgene in order to determine if over-expression of *CECR2* and/or *SLC25A18* would produce any CES related phenotypes. The human BAC contained the *CECR2* and *SLC25A18* genes in their entirety, and in a preliminary experiment *CECR2* appeared to be expressed. Two

transgenic lines were generated and bred. Line 1 showed no abnormalities in any animals, even when transgenic animals were mated together in an attempt to increase the transgene copy number. Line 10, however, proved difficult to maintain, as females would often appear pregnant and then fail to give birth. Even though we were propagating this line by mating to normal FVB/n animals, there was clearly some *in utero* lethality occurring, as cannibalization was ruled out based on a lack of pup remains/blood in cages. It was not clear, however, if this lethality was due to expression of genes from the BAC transgene, was due to a dominant insertional mutagenesis effect, or was completely unrelated to the transgene. If the lethality was associated with a dominant insertional mutagenesis effect (in FBV/n), we might expect to see the same lethality in the transgene rescue experiment (see **4.5**), because the effect should be independent of *Cecr2* genotype. It is possible, however, that there are modifier loci in 129P2 and BALB/c, which mute this effect. We did not observe transgene-associated lethality in the rescue experiment, which would argue against an insertional mutagenesis effect. If the lethality was associated with expression of a gene(s) from the BAC, then it would appear that the transgene in Line 1 is not functional as no lethality was observed in that line, or that the transgene copy number was not high enough to produce an effect. It is possible however, that the mixed background in the rescue experiment may reduce penetrance of this lethality. Unfortunately, this experiment is still in progress and this question cannot be answered at this time. Once the transgenic Line 10 (FVB/n) is no longer in danger of extinction, it would be useful to mate two Line 10 transgenics together and harvest embryos in an attempt to determine the basis of the lethality effect.

The fact that the transgene has not been thoroughly characterized, does somewhat limit the possible interpretations of this rescue experiment. We cannot be certain that all regulatory elements required for *CECR2* expression are contained within BAC 357F7. Certainly, any elements more than 35 kbp centromeric to *CECR2* exon 1 would not be included on the BAC. Mapping of such elements has not been performed, but would be of interest in determining whether specific *CECR2* transcripts are temporally or spatially regulated. Considering that regions of homology between human and mouse lie within *CECR2/Cecr2* intron 1, this region would be a logical area in which to initiate such an analysis.

In order to determine potential causes of *in utero* death of the Line 10 embryos, various approaches could be taken. Recently, non-invasive techniques have been utilized to visualize mouse embryonic development. Magnetic resonance imaging (MRI) has been used to track mouse development over a time course (Chapon *et al.* 2002). Such a technique has limitations in resolution, but would allow embryonic development to be tracked without having to sacrifice the embryos. Doppler echocardiography (ECG) has also been used on mouse embryos to study heart function in vivo. Transcriptional coactivator *Cited2* knockout mice display neural tube defects and tetrology of Fallot, the latter of which has been shown *in utero* with Doppler ECG (Yin *et al.* 2002). ECG of *CECR2* transgenic mice could potentially determine if a congenital heart defect is present. However, such experiments are not warranted until expression of, and protein production from, the *CECR2* transgene is confirmed.

4.5 Human *CECR2* containing BAC does not rescue *Cecr2* mutants

We also crossed the BAC 357F7 transgene (Lines 1 and 10) into the $Cecr2^{+/pGT1}$ mice in order to determine if the presence of *CECR2* would rescue $Cecr2^{pGT1/pGT1}$ animals from exencephaly. Due to time and space constraints, Line 1 was not tested beyond a few initial matings and, therefore, will require further study. We focused on Line 10 because there appeared to be some lethality in the FVB/n transgenic animals, although it was not clear that the *in utero* lethality was related to the BAC transgene. In the Line 10 rescue experiment, the results varied depending upon which stud male was used. There are several possibilities as to why this may have occurred. Considering that each stud male was produced from an intercross between a 129P2/Ola:BALB/c animal ($Cecr2^{+/pGT1}$) and a FVB/n animal ($Cecr2^{+/+}$, +TgN(*CECR2*)), their genetic makeup would not be identical. We do know that there are modifier loci in FVB/n that reduce penetrance of exencephaly caused by the *Cecr2* mutation. Different modifier loci may have been inherited in each of the different stud mice, thereby causing a different penetrance when backcrossed to 129P2:BALB/C *Cecr2* heterozygotes. Conversely, several transgene insertions may have occurred during the initial pronuclear injection of the BAC transgene, which could also explain the variation in the transgene rescue experiment results. If there were multiple insertions of the transgene, with some being functional and others not, they would likely segregate from one another, producing sub-lines containing either a functional copy or a non-functional copy of the BAC transgene, which would produce contrasting experimental results. Considering that the PCR tracking assay does not differentiate between these possibilities, the only way of identifying a functional transgene from a non-functional one would be to either perform a rescue experiment or to test for CECR2

expression with a CECR2 specific antibody, in addition to *in situ* hybridization of mouse chromosomes to ensure only one transgene array is present. The rescue experiment clearly showed that the two individual Line 10 animals did not carry a transgene capable of rescuing the *Cecr2* mutation. Currently, we are unable to determine if this is because the transgene is not producing a functional protein, or whether said protein is not functional in mouse. The rescue experiment will ultimately benefit from being performed on congenic backgrounds of each of the three strains used in this study (129P2/Ola, BALB/c, and FVB/n).

4.6 *Cecr2* is potentially important for neurulation

Altering the levels of Cecr2 protein in the developing mouse embryo can have catastrophic effects, which are incompatible with life. From our $Cecr2^{pGT1/pGT1}$ mice we have been able to determine that *Cecr2* is expressed in tissues involved in neurulation, in addition to other, non-neural tissues in the embryo. Based on the exencephaly phenotype of *Cecr2* mutant embryos, we can infer that Cecr2 is likely important for neural tube closure. What is not clear, however, the quantity of Cecr2 that is required to stimulate neural tube closure. Because our $Cecr2^{pGT1/pGT1}$ mice are not complete null mutations, some undetermined amount of normal Cecr2 is produced as a result of read-through and bypass splicing of the genetrap. We are currently unable to determine the quantity of normal *Cecr2* mRNA and Cecr2 protein that is produced in mutant mice. To determine mRNA levels, either quantitative RT-PCR or RNase protection will be required, as Northern blotting and *in situ* hybridization are not effective on *Cecr2*. To determine protein levels in Cecr2 mutant mice, Cecr2 specific antibodies will be required, and these

are currently under development. It should be noted, however, that due to the high homology between CECR2 and Cecr2, it is unlikely that we will be able to produce human or mouse specific antibodies for a significant portion of the protein. The only region of CECR2/Cecr2 in which species specific antibodies seem plausible is within the carboxy-half of the protein. Another, somewhat less likely possibility for the exencephaly phenotype of *Cecr2*$^{pGT1/pGT1}$ mice, is that the pGT1 allele is a gain-of-function mutation, and the exencephaly phenotype is the result of a novel function of the Cecr2-βgeo fusion protein. However, gain-of-function mutations, such as the tri-nucleotide repeat expansion class of diseases (see Huntington disease MIM# 143100 and Spinocerebellar ataxia 1 MIM# 164400) are typically dominant in inheritance, whereas the *Cecr2*$^{pGT1/pGT1}$ allele appears to be recessive in nature, based on our observations of exencephaly penetrance. This would argue for *Cecr2*$^{pGT1/pGT1}$ being a loss-of-function allele.

I feel confident that the exencephaly phenotype is directly correlated to *Cecr2* genotype, as the genetrap has been passed through 4 generations with no change in the genotype/exencephaly ratio. A plausible way that the exencephaly could be unrelated to *Cecr2* genotype would be if a second genetrap insertion was in a gene tightly linked to *Cecr2*, as we have never seen exencephaly in animals with a *Cecr2*$^{+/+}$ genotype. It would be prudent to confirm that only one genetrap insertion is present in our ES genetrap cell line by Southern blotting. A position effect on a nearby gene is also plausible, although the distance to the next gene(s) is large; ~50 kbp downstream and >100 kbp upstream, and the small amount of inserted sequence (~ 6 kbp) makes this possibility less likely. Characterization of the expression levels of neighbouring genes would, however, address this possibility.

Creation of a second chimera from the same ES cell line that was used to produce our knockout animal does not seem necessary, once it is confirmed that the genetrap insertion is only present within *Cecr2*. Creation of a new allele would be more appropriate in confirming our results. Generation of a second mouse from a genetrap in *Cecr2* exon 1 is currently underway. If exencephaly results from this allele, which would lack all of the recognizable amino acid motifs in Cecr2, we should be confident that the exencephaly phenotype results from loss of Cecr2 function. In order to determine if Cecr2 is essential for neural tube closure (in 100% of animals), a null allele would be required (see *4.7.2iii* for more discussion). There are few examples of mutations in which 100% exencephaly penetrance is observed. The *Cart1* homeobox gene is one such example, but only on a 129Sv background (Zhao *et al.* 1996). *Cart1* has reduced penetrance (65%) on a C57BL/129 hybrid background, suggesting at least one modifier locus in C57BL/6J is present (Zhao *et al.* 1996). This is similar to what we experienced with strain FVB/n and the *Cecr2* mutation. The fact that mutations in most NTD genes are not 100% penetrant is likely a result of two factors: (1) there may be redundancy built into the system over evolution such that humans are less susceptible to NTDs in the event of a single gene mutation (Brook *et al.* 1991; van Straaten *et al.* 1993), and (2) NTDs likely manifest due to threshold effects based on a variable "liability" in the population, which is comprised of environmental and genetic factors, making it multifactorial (Fraser 1976; Colas and Schoenwolf 2001). Once beyond a certain threshold level, NTDs will manifest. The fact that various unmapped loci have an effect on exencephaly penetrance in different NTD mutants suggests that it is likely that multiple loci genetically interact during the neurulation process. These loci could, then, theoretically act as susceptibility

factors for NTDs in mice and humans when in a mutant form. We would not expect to see several mutant loci together in common inbred laboratory mouse strains. There is, however, one mouse strain that illustrates a multifactorial inheritance of exencephaly. The SELH/Bc strain likely has 2 or 3 major loci contributing to its 30% rate of exencephaly (Juriloff *et al.* 2001). The SELH/Bc strain is likely more analogous to humans in which multiple genes act as susceptibility factors for NTDs. To date, no human genes have been identified that directly contribute to anencephaly. This may be due to redundancy built into the human system, or could simply be due to the fact that investigators have mostly been looking at spina bifida cases because they are relatively easy to obtain due to viability. We would not necessarily expect these spina bifida cases to be caused by mutations in anencephaly causing genes. In a recent study, only 5% of the samples tested for mutations in the exencephaly causing NTD genes *Tfap2α* and *Msx2* were anencephalic, with the vast majority being spina bifida (Stegmann *et al.* 2001). Other studies have been similar in their use of spina bifida samples (Trembath *et al.* 1999; Morrison *et al.* 1998) and searching for mutations in seemingly random NTD genes. The relative genetic diversity of humans may make the presence of particular combinations of modifier loci a much more rare event than in laboratory mice. It would certainly be interesting to determine if *CECR2* has a role in neurulation in humans. This would require collection of anencephalic embryo abortuses, as most anencephaly cases are detected with prenatal testing and subsequently aborted, in a study to look for *CECR2* mutations. However, given the large size of *CECR2*, this task would be highly labour intensive.

4.7 Folic acid cannot rescue *Cecr2* mutants

Ideally one would like to perform a maternal folic acid supplementation experiment on a congenic background. Unfortunately, that option was not available to us, so we performed the experiment on a mixed 129P2:BALB/c background. This small experiment indicated that intraperitoneal injection of folic acid provided no protection from the effects of the *Cecr2* mutation. In other examples of folic acid rescue, NTD penetrance can drop dramatically, or only slightly, depending on the mutation. For example, *Cart1* knockouts exhibit 100% penetrance in control animals, but only 20% in folic acid treated animals (Zhao *et al.* 1996). In contrast, penetrance in the *Cd* (Crooked) mutation is only reduced from 20% to 15% (Carter *et al.* 1999). While we cannot be certain of the actual plasma folate levels in the injected females, this approach has been successfully used in several mouse studies with success (Zhao *et al.* 1996; Barbera *et al.* 2002). Culturing embryos in folate-supplemented medium has also been used successfully (Fleming and Copp 1998). The embryo culturing method is more technically demanding, but allows for more accurate determination of folate levels.

Furthermore, other mutants do not respond to folic acid at all. *Axd* (axial defects) and *ct* (curly tail) are not responsive to folic acid, but do show a decrease in NTD frequency when supplemented with methionine or inositol, respectively (Juriloff and Harris 2000). It remains to be seen as to whether or not the *Cecr2* induced NTDs are sensitive to these other compounds.

It has been suggested that folic acid is required for timed proliferative bursts during human development (Antony and Hansen 2000). Folic acid may potentially rescue defects that slow these bursts, such as a NTD gene mutation. Folate metabolism and

NTDs have been extensively studied. They have been linked in various studies (see *1.5.3*) in several ways. However, to date there has not been a convincing argument as to why folate can rescue some mouse NTDs and not others. Folate is an important molecule in several cellular metabolic cycles, including purine biosynthesis, methionine re-methylation from homocysteine, histidine metabolism, and formate metabolism (reviewed in van der Put *et al.* 2001). Folate is also indirectly implicated in DNA methylation, and subsequent gene expression, via S-adenosylmethionine as a major cellular methyl group donor (van der Put *et al.* 2001). If folate metabolism is altered, the ratio of S-adenosylmethionine to S-adenosyhomocysteine may be altered leading to a global shortage of methyl groups and subsequently global demethylation (Jhaveri *et al.* 2001). This would theoretically allow up-regulation of genes that are normally repressed via methylation. Global demethylation has been shown via analysis of folate depleted versus folate repleted cells, in addition to microarray analysis showing that the expression levels of 8 of 2200 genes tested are altered in folate depleted cells (Jhaveri *et al.* 2001). Extrapolated to 30,000 genes in the mammalian genome, this would suggest that expression levels of approximately 110 genes could be altered under low folate conditions. Whether or not these hypothetical genes could be implicated in NTDs remains to be determined. However, it is an attractive theory that folate rescue of NTDs may be due to its effect on the expression of particular genes. Such genes could genetically interact with a particular mutant locus; altering expression levels of the downstream gene may potentially 'suppress' the mutant allele. Whether such a model is viable remains to be seen, but would require a whole genome microarray analysis in various mouse mutants in conjunction with folate depleted and repleted status.

Previous speculation has suggested that elevated levels of homocysteine, due to defects in folate metabolism, can induce NTDs directly, as studies have shown that mothers of children with NTDs showed elevated levels of homocysteine (Steegers-Theunissen *et al.* 1994; Mills *et al.* 1995). Polymorphisms in the *MTHFR* gene, which is involved in the production of 5-methylenetetrahydrofolate, have been implicated as a risk factor in NTDs (van der Put *et al.* 1995; van der Put 1996), possibly via an increased level of homocysteine. However, animal models have produced conflicting results regarding homocysteine and NTDs. Homocysteine apparently has no effect on neurulation in rat (van Aerts *et al.* 1994) or mouse embryos (Greene *et al.* 2003), but induces NTDs in chick embryos (Rosenquist *et al.* 1996). In another case, the curly-tail (*ct*) mouse displays altered homocysteine metabolism in the absence of a *Mthfr* defect, and is unresponsive to folate (Tran *et al.* 2002).

Until recently it was unclear as to why folic acid and homocysteine had apparently contrasting effects on neurulation. Recently is was shown that both folic acid and homocysteine have effects on neural crest and neuroepethelial growth and differentiation in neural tube explants *in vitro*. Folic acid increases outgrowth of neuroepithelial cells and induces the differentiation of neural crest into nerve and muscle, while homocysteine inhibits neural crest differentiation and increases neural crest migration away from the neural tube (Boot *et al.* 2003). This leads to a shortage of neuroepithelial cells within the neural tube during neural tube closure. However, this cannot explain all NTDs, as all NTD genes do not affect folate metabolism. Only *Folbp1* (Piedrahita *et al.* 1999) and *Pax3* (Fleming and Copp 1998) mouse mutants have been shown to have defects in folate metabolism. Other folate insensitive mouse mutants may

be due to altered homocysteine metabolism or inadequate S-adenosylmethionine leading to reduced DNA methylation and gene expression (van der Put *et al.* 2001). Testing of the *Cecr2* mutant to determine if folate or homocysteine metabolism is altered, or if global DNA methylation levels are altered, may provide clues to the specific nature of the *Cecr2* mutation. We predict, however, that there will not be a folate metabolism defect in *Cecr2* mice if they are not rescued by folate supplementation.

4.8 Cat eye syndrome: monogenic or polygenic?

Further study of the *CECR2* transgenic mice and the *Cecr2* knockout mouse in congenic lines is required and may shed light on possible *CECR2* involvement in CES. However, the preliminary evidence presented in this thesis has not produced further evidence for, or against, a prominent role for *CECR2* in CES. Thus, we need to ask ourselves whether there is one "major" gene producing CES or is the syndrome the result of the synergy of several over-expressed genes? This question is a difficult one to answer, as both possibilities are equally plausible.

Considering the embryonic basis of CES affected tissues, there is no common embryonic origin that might suggest a single gene mutation may be causing the pleiotrophic CES phenotype. In CES, affected tissues include the iris, kidney, heart vessels, outer ear and anus (Shinzel *et al.* 1981). The iris portion of the eye is derived from ectodermal outgrowths from the brain, kidney is derived from intermediate mesoderm, the epithelium of the anus is derived from ectoderm, the outer ear is derived from the first branchial groove and its associated mesoderm, and the heart vessels are derived from aortic arches which are comprised of head mesenchyme (with essential

contributions from neural crest) (Moore and Persaud 1993; Kirby and Waldo 1995). None of the CES genes studied to date have shown expression in all of these embryonic tissues. This may be suggestive of a polygenic origin for CES.

There are, however, examples of disorders with pleiotrophic phenotypes, like CES does, that fall into both categories. For example, mutations in the transcriptional coactivator *CBP* alone cause Rubinstein-Taybi syndrome, which has various phenotypes, including heart defects, hand/foot abnormalities, and mental retardation (Petrij *et al.* 1995). In contrast, Williams syndrome is caused by a 1.5 Mbp deletion of which certain phenotypes can be ascribed to the lack of particular genes. For example, the heart defect (supravalvular aortic stenosis) in Willams syndrome can be correlated with haploinsufficiency of the Elastin (*ELN*) gene and this has been confirmed by independent mutations in the gene (Morris and Mervis 2000, and references therein). Yet the behavioural phenotypes of Williams syndrome are not correlated to the *ELN* gene and, therefore, must be the result of the loss of one or more of the remaining genes within the region. It is plausible to imagine synergistic effects of multiple genes when they are over or under-expressed. At this point, it is not clear whether CES is the result of the over-expression of one or more genes. Mouse modeling is our method of choice in order to analyze these possibilities, yet to date has not yielded any abnormal phenotype resulting from the over-expression of human genes. This approach has not been exhausted, however, as there are several more genes in the CESCR that remain to be tested in this method, notably *CECR6*. The possibility does exist that multiple genes must be overexpressed together to produce any abnormal phenotype. This would be addressed most effectively by duplicating the entire region in mouse, which is a technically

demanding experiment. A 22 cM duplication has been generated with the cre-lox system (Zheng *et al.* 2000) demonstrating the feasibility of the method. A simpler approach may be to cross single transgenics together, although this would make tracking the transgenes much more complicated. In addition, the fact that *CECR1* is not present in mouse means that an engineered duplication of mouse would not be a faithful recreation of a human CES duplication. As such, it will be impossible to create the perfect mouse model of a CES-like duplication.

4.9 Future directions

4.9.1 Characterization of the CECR2/Cecr2 protein

i) Interacting proteins

Identification of CECR2 interacting proteins and determining whether they are implicated in NTDs will help us to make connections between genes and to better understand neurulation pathways. To date, all information about possible function of CECR2 has been implied based upon sequence motifs. These motifs suggest that CECR2 is likely a transcriptional regulator and, as such, we would expect CECR2 to be localized to DNA and interact with numerous other proteins. Truncated forms of Cecr2 fused to β-galactosidase have been shown to be nuclear localized and DNA associated (Tate *et. al.* 1998), but full length CECR2/Cecr2 localization has not been shown. CECR2/Cecr2 antibodies will be required to confirm nuclear localization by immunofluorescence. Antibodies will also be useful in determining interacting proteins, via co-immunoprecipitation and affinity chromatography. Two-hybrid screens may be potentially fraught with false positives, considering that CECR2 may be a transcriptional

activator and could potentially activate reporter gene expression on its own. For this reason, co-immunoprecipitation or affinity chromatography and subsequent protein identification by mass spectrometry would seem to be a more sensible approach.

ii) DNA binding experiments

Currently, the only evidence that suggests that CECR2 directly interacts with DNA is the putative AT hook motif. DNA binding experiments are required to determine if CECR2 does bind DNA, what part of the protein is necessary for DNA binding, and what the binding site consensus sequence is. This would be performed in a manner similar to that used by Swan *et al.* 2001, in which the recombinant versions of the *Theileria annulata* proteins TashAT1 and TashAT2 were incubated with random, labeled double stranded oligonucleotides. After several rounds of binding/purification, the oligonucleotides were PCR amplified and sequenced to identify the sequences of those oligos binding to the recombinant protein. If a consensus sequence was identified this would open the door to a bioinformatics project looking for potential target genes of *CECR2* activation, assuming that *CECR2* is gene activator. Identification of potential target genes would allow characterization of their expression in *Cecr2* mutants. *CECR2* target genes may form part of a pathway involved in neurulation.

4.9.2 Further characterization of Cecr2 mutants

i) Generation of a Cecr2 null allele

The pGT1 genetrap allele of *Cecr2* is not a null allele, as some undetermined amount of normal transcript is produced based on RT-PCR experiments. The amount and

type of the remaining *Cecr2* isoforms needs to be clarified. Furthermore, it will be of interest to determine the exencephaly penetrance of a *Cecr2* null allele. Is the ~66% exencephaly penetrance of the pGT1 allele on a 129P2/BALB background simply based on stochastic factors and threshold effects, or is it limited by the background of normal *Cecr2* transcripts produced? Due to the reduced penetrance of many NTD genes, based on putative threshold effects, we should not necessarily assume that a *Cecr2* null allele would produce 100% exencephaly penetrance. Other, possibly lethal phenotypes may be produced by a null allele, which we did not observe due to our hypomorphic *Cecr2* allele.

In addition, other, lower penetrance, phenotypes may be observed in *Cecr2* null alleles in addition to possible strain specific phenotypes. An example of variable penetrance and expressivity is found with mice null for *Ski*, a proto-oncogene. As the *Ski* mutation is moved from a 129 background to a C57BL background, exencephaly drops from 83% to 5%, while facial clefting increases from 0% to 93% penetrance (Colmenares *et al.* 2001). Clearly, different mouse strains contain various modifier loci that can influence different aspects of development in conjunction with NTD genes.

ii) Tunel staining to observe apoptosis patterns

Other transcriptional co-activator and transcription factor mouse mutants have been shown to display abnormal apoptosis patterns during development. *Cited2* (Bamforth *et al.* 2001), *Pax3* (Phelan *et al.* 1997), and *Ap2* (Schorle *et al.* 1996) mutants all have increased levels of apoptosis in the midbrain region, as shown by Tunel staining, in addition to NTDs. It is not clear if this is a primary result of the mutation or a secondary result of an altered developmental pattern. Recently, some evidence has been

presented that directly links apoptosis and neurulation. *Pax3* mutants can be rescued by *p53* loss of function mutations, suggesting that loss of *p53* dependant apoptosis prevents neural tube closure and that *Pax3* works to inhibit *p53* dependant apoptosis (Pani *et al.* 2002). This suggests an intriguing possibility, that apoptosis may be inhibited by some NTD causing genes, such as those NTD genes involved in gene regulation. This is in apparent opposition to the theory that NTD genes positively act to cause neural tube closure. However, reduced apoptosis has also been observed in NTD mutants. *Jnk1/Jnk2* double mutants display reduced apoptosis in the hindbrain but conversely, show increased apoptosis in the forebrain (Sabapathy *et al.* 1999). It is still too early to tell if NTD genes are directly involved in apoptotic pathways, or that the inappropriate apoptosis observed in NTDs is due to an altered developmental pathway.

It would be prudent to test apoptosis patterns in *Cecr2* mutants to ask if the *Cecr2* mutation results in inappropriate apoptosis during neurulation. If $Cecr2^{pGT1/pGT1}$ mice do show increased levels of apoptosis, a cross to $p53^{-/-}$ mice could be performed to determine if $Cecr2^{pGT1/pGT1}$ induced exencephaly can be rescued.

iii) Folic acid rescue experiment

Despite the fact that we showed that folic acid likely does not affect the *Cecr2* induced exencephaly, this experiment was performed on a mixed background containing 3 strains. The experiment should be repeated once the *Cecr2* mutation has been moved onto each of the 3 strains (BALB/c, 129P2/Ola and FVB/n) for at least 5 generations, to ensure relative purity of the strain background as per guidelines set out by Jackson

Laboratories (www.jaxmice.jax.org). This will give a more accurate representation as to the effects of maternal folic acid supplementation in *Cecr2* mutants.

The results of a folic acid experiment will be interesting in light of the folic acid rescue status of other transcriptional regulators when compared to inappropriate apoptosis. For example, *Cited2*, a transcriptional co-activator that displays exencephaly and abnormal apoptosis when knocked out, is rescued by maternal folic acid supplementation (Bamforth et. al 2001; Barbera *et al.* 2002). *Pax3* is another example of a gene that causes inappropriate apoptosis and NTDs when mutated, but is partially rescued by folic acid (Phelan et. al 1997; Fleming and Copp 1998). Transcription factor *Ap2* has also been shown to cause exencephaly and increased apoptosis when mutated, but has not yet been tested for folic acid rescue (Schorle *et al.* 1996). The relationship between folates and inappropriate apoptosis is an intriguing one and begs the question as to the relationship with neurulation. This is an area of active research, but no attractive models yet exist.

iv) CECR2 transgene rescue experiment

Our original *CECR2* transgene rescue experiment was somewhat problematic insofar as the FVB/n strain clearly muted the *Cecr2* mutant phenotype. In addition, the presence of 3 strain backgrounds in the mice allowed for different modifier loci to be present in different individuals. This experiment would benefit from pure strain backgrounds. Therefore, the *Cecr2* mutation and the *CECR2* transgene are being backcrossed together onto each of the three different strain backgrounds (BALB/c, 129P2/Ola and FVB/n). After a minimum of five generations, the experiment will be

repeated by intercrossing *Cecr2* heterozygotes, one of which carries the *CECR2* transgene. At the same time, *Cecr2* heterozygotes without the transgene will be intercrossed to determine exencephaly penetrance and also to look for any other potential strain specific abnormalities. Prior to undertaking this experiment, it would seem prudent to perform a detailed analysis of the *CECR2* transgene in order to determine if it is expressed and producing protein.

v) Comparative gene expression analysis between Cecr2 genotypes

Given the high probability that *CECR2/Cecr2* is involved in the expression of other genes, it would be of interest to determine which genes' expression are affected when *Cecr2* is removed. Comparative cDNA hybridization between wildtype and $Cecr2^{pGT1/pGT1}$ samples from embryos undergoing neurulation could potentially yield information regarding *Cecr2* target genes. This would involve high-density gene specific oligonucleotides arrays (Chips) hybridized with labeled cDNA generated from the neural tube of different mouse samples. This technique has been successful in profiling gene expression patterns in the developing neural tube of mice (Finnell *et al.* 2002). While microarray analysis is fraught with technical difficulties, the potential to identify numerous target genes at the same time cannot be overlooked. This information could certainly be useful in identifying gene hierarchies when building neurulation pathways.

vi) Generation of double mutants with other NTD genes

In order to elucidate the relationship between different exencephaly causing genes, crossing the different mutations may be informative. If double heterozygotes were

generated between two different exencephaly causing genes and the result was a synergistic effect that increased the rates of exencephaly, we might assume that the genes operate in an analogous pathway. Synergistic mutations could also show epistasis. The generation of double mutants may turn out to be a powerful tool in dissecting the relationship between the different NTD genes. The precedence for *p53* mutations suppressing the *Pax3* mutation is evidence that this approach is viable and potentially informative (Pani *et al.* 2002). Another double mutant example includes the identification of a genetic interaction between *undulated* and *Patch* in which double mutant mice display spina bifida while neither single mutant does (Helwig *et al.* 1995). *Patch* mice are due to deletion of the platelet derived growth factor alpha gene (*Pdgfa*) (Stephenson *et al.* 1991), while *undulated* mice are either due to a point mutation (Balling *et al.* 1988) or a deletion of the last exon of *Pax1* (Chalepakis *et al.* 1991). It has subsequently been shown that the transcription factor *Pax1* is involved in the regulation of expression of *Pdgfa* (Joosten *et al.* 1998). If *Cecr2* is involved in the regulation of expression of other genes, it may show epistasis with them. In another interesting example of epistasis, it has been demonstrated that double heterozygotes of *Nf1* and *Pax3* exhibit a low penetrance of NTDs, while neither single heterozygote alone does (Lakkis *et al.* 1999), indicating that *Nf1* is a modifier of the *Pax3* locus. Studies like these may yield important information about genetic interactions between NTD genes and help us to build neurulation pathways.

4.9.3 Generation of a Drosophila mutant

Currently, the only known homolog of *CECR2* identified in a genetically manipulatable organism other than mouse is the *Drosophila* gene *CG10115*. However, no

known mutant alleles of *CG10115* exist. We did, however, obtain a strain containing a P-element insertion within intron 1 of *CG10115*. We performed excision of the P-element in an attempt to generate a deletion surrounding the insertion site, potentially disrupting the *CG10115* transcript. We were unsuccessful in this regard, yet this avenue has not been exhausted. This experiment can be modified by adding markers proximal and distal to the *CG10115* locus such that cross over events can be detected that can potentially generate deletions at the *CG10115* locus. This series of experiments should not be abandoned as the genetic power of *Drosophila* has the potential to uncover much information about *CG10115/CECR2*. If a deletion in *CG10115* produces a recognizable phenotype, suppressor screens may allow the identification of genetically interacting proteins, which would certainly complement the physical interaction studies in human and mouse. With the ultimate goal being to generate the genetic pathway(s) leading to neural tube closure, these potential interactions will allow us to add several pieces to the apparently large puzzle. Considering that over 70 genes have been implicated in NTDs in mice, it would seem that multiple pathways are likely, in addition to redundancy built into the system. Using a model system like mouse to dissect these pathways is technically feasible, yet financially and temporally difficult. For this reason, using a system like *Drosophila* to identify potential pathway members could allow a much more targeted approach than in the mouse system.

4.10 Concluding remarks

The work contained within this document has focused on the gene *CECR2* and its mouse ortholog *Cecr2*. Prior to this study, these genes had not been identified. This work

presents a preliminary characterization of both genes, and potentially implicates *Cecr2* in neurulation. *Cecr2* joins a rather large list of neurulation genes in mouse, but is atypical due to the fact that exencephaly is the only defect detected in mouse mutants. This is of interest because it potentially makes *Cecr2* a good mouse model of NTDs, without other, complicating, phenotypes. The *Cecr2* mouse should provide a useful system for the further study of neurulation pathways in mouse. The initial impetus for studying *CECR2* was that it is found in the cat eye syndrome critical region. We have, however, been unable to exclude *CECR2* as a CES candidate gene. Further study of *CECR2* is, therefore, required.

REFERENCES

Aihara T, Miyoshi Y, Koyama K, Suzuki M, Takahashi E, Monden M, Nakamura Y (1998) Cloning and mapping of SMARCA5 encoding hSNF2H, a novel human homologue of Drosophila ISWI. *Cytogenet Cell Genet* 81:191-193

Akimaru H, Chen Y, Dai P, Hou DX, Nonaka M, Smolik SM, Armstrong S, Goodman RH, Ishii S (1997a) Drosophila CBP is a co-activator of cubitus interruptus in hedgehog signalling. *Nature* 386:735-738.

Akimaru H, Hou DX, Ishii S (1997b) Drosophila CBP is required for dorsal-dependent twist gene expression. *Nat Genet* 17:211-214.

Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA* 51:786-794.

Altafaj X, Dierssen M, Baamonde C, Marti E, Visa J, Guimera J, Oset M, Gonzalez JR, Florez J, Fillat C, Estivill X (2001) Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. *Hum Mol Genet* 10:1915-1923.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 21:403-410.

Antonsson B (1997) Phosphatidyinositol synthase from mammalian tissues. *Biochim Biophys Acta* 1348:179-186.

Antony AC, Hansen DK (2000) Hypothesis: folate-responsive neural tube defects and neurocristopathies. *Teratology* 62:42-50.

Aravind L, Landsman D (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* 26:4413-4421.

Armstrong JF, Kaufman MH, Harrison DJ, Clarke AR (1995) High-frequency developmental abnormalities in p53-deficient mice. *Curr Biol* 5:931-936.

Attwood TK, Blythe M, Flower DR, Gaulton A, Mabey JE, Maudling N, McGregor L, Mitchell A, Moulton G, Paine K, Scordis P (2002) PRINTS and PRINTS-S shed light on protein ancestry. *Nucl Acids Res* 30:239-241.

Balling R, Deutsch U, Gruss P (1988) undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1. *Cell* 155:531-535.

Bamforth SD, Braganca J, Eloranta JJ, Murdoch JN, Marques FI, Kranc KR, Farza H, Henderson DJ, Hurst HC, Bhattacharya S (2001) Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. *Nat Genet* 29:469-474.

Bannister AJ, Kouzarides T (1996) The CBP co-activator is a histone acetyltransferase. *Nature* 284:641-643.

Barbera JP, Rodriguez TA, Greene ND, Weninger WJ, Simeone A, Copp AJ, Beddington RS, Dunwoodie S (2002) Folic acid prevents exencephaly in Cited2 deficient mice. *Hum Mol Genet* 11:283-293.

Barlev NA, Liu L, Chehab NH, Mansfield K, Harris KG, Halazonetis TD, Berger SL (2001) Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol Cell* 8:1243-1254.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer E.L (2002) The Pfam Protein Families Database *Nucl Acids Res* 30:276-280.

Baud V, Mears AJ, Lamour V, Scamps C, Duncan AMV, McDermid HE, Lipinski M (1994) The E subunit of vacuolar H^+-ATPase localizes close to the centromere on human chromosome 22. *Hum Mol Genet* 3:335-339.

Blake KD, Davenport SLH, Hall BD, Hefner MA, Pagon RA, Williams MS, Lin AE, Graham JM Jr. (1998) CHARGE association: an update and review for the primary pediatrician. *Clin Pediatr* 37(3):159-173.

Boot MJ, Steegers-Theunissen RP, Poelmann RE, Van Iperen L, Lindemans J, Gittenberger-De Groot AC (2003) Folic acid and homocysteine affect neural crest and neuroepithelial cell outgrowth and differentiation in vitro. *Dev Dyn* 227:301-308.

Bridgland L, Footz TK, Kardel MD, Riazi MA, McDermid HE (2003) Three duplicons form a novel transcription unit in the pericentromeric region of chromosome 22q11. *Hum Genet* 112:57-61.

Brook FA, Shum AS, Van Straaten HW, Copp AJ (1991) Curvature of the caudal region is responsible for failure of neural tube closure in the curly tail (ct) mouse embryo. *Development* 113:671-678.

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD. (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84:843-851.

Bucher P., Bairoch A. (1994) A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation In "ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology". (Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds.), pp53-61, AAAI Press, Menlo Park.

Burge C and Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78-94.

Burn J and Goodship J (1996) Congenital heart disease. In: Rimoin DL, Connor JM, Pyeritz RE (eds) Emery and Rimoins principles and practice of medical genetics, 3d ed. Vol 1:767-828.

Byers, PH (1993) Osteogenesis imperfecta. In: Royce, P. M.; Steinmann, B. : Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects. New York: Wiley-Liss (pub.) Pp. 317-350.

Carter M, Ulrich S, Oofuji Y, Williams DA, Ross ME (1999) Crooked tail (Cd) models human folate-responsive neural tube defects. *Hum Mol Genet* 8:2199-2204.

Chance PF, Alderson MK, Leppig KA, Lensch MW, Matsunami N, Smith B, Swanson PD, Odelberg SJ, Disteche CM, Bird TD (1993) DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* 72:143-151.

Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR (1999a) Regulation of transcription by a protein methyltransferase. *Science* 284:2174-2177.

Chen H, Lin RJ, Xie W, Wilpitz D, Evans RM (1999b) Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell* 98:675-686.

Chen ZF, Behringer RR (1995) twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev* 9:686-699.

Chalepakis G, Fritsch R, Fickenscher H, Deutsch U, Goulding M, Gruss P (1991) The molecular basis of the undulated/Pax-1 mutation. *Cell* 66:873-884.

Chapon C, Franconi F, Roux J, Marescaux L, Le Jeune JJ, Lemaire L (2002) In utero time-course assessment of mouse embryo development using high resolution magnetic resonance imaging. *Anat Embryol* (Berl). 206:131-7.

Colas JF and Schoenwolf GC (2001) Towards a cellular and molecular understanding of neurulation. *Dev Dyn* 221:117-145.

Collier B, Goobar-Larsson L, Sokolowski M, Schwartz S (1998) Translational inhibition in vitro of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogenous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2. *J Biol Chem* 273:22648-22656.

De Arcangelis A, Mark M, Kreidberg J, Sorokin L, Georges-Labouesse E (1999) Synergistic activities of alpha3 and alpha6 integrins are required during apical ectodermal ridge formation and organogenesis in the mouse. *Development* 126:3957-3968.

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399:491-496.

Dorer DR (1997) Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res* 6:3-10.

Dressler GR, Wilkinson JE, Rothenpieler UW, Patterson LT, Williams-Simons L, Westphal H (1993) Deregulation of Pax-2 expression in transgenic mice generates severe kidney abnormalities. *Nature* 362:65-67.

Dunham I *et al.* (1999) The DNA sequence of human chromosome 22. *Nature* 402:489-495.

Dynlacht BD, Hoey T, Tjian R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66:563-576.

Dyck PJ, Thomas PK, Griffin JW, Low PA, Poduslo JF (1993) *Peripheral neuropathy*, volume 2 (W.B. Saunders Company, Philadelphia, PA).

Eccles MR and LA Schimmenti (1999) Renal-coloboma syndrome: a multi-system developmental disorder caused by *PAX2* mutations. *Clin Genet* 56:1-9.

Edelmann L, Pandita RK, Morrow BE (1999) Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet* 64:1076-1086.

el Ghouzzi V, Le Merrer M, Perrin-Schmitt F, Lajeunie E, Benit P, Renier D, Bourgeois P, Bolcato-Bellemin AL, Munnich A, Bonaventure J (1997) Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat Genet* 15:36-41.

Essien FB, Wannberg SL (1993) Methionine but not folinic acid or vitamin B-12 alters the frequency of neural tube defects in Axd mutant mice. *J Nutrition* 123:27-34.

Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A (2002) The PROSITE database, its status in 2002. *Nucl Acids Res* 30:235-238.

Fan XC, Steitz JA (1998a) HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proc Natl Acad Sci U S A* 95:15293-15298.

Fan XC, Steitz JA (1998b) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J* 17:3448-3460.

Favor J, Sandulache R, Neuhauser-Klaus A, Pretsch W, Chatterjee B, Senft E, Wurst W, Blanquet V, Grimes P, Sporle R, Schughart K (1996) The mouse Pax2(1Neu) mutation is identical to a human PAX2 mutation in a family with renal-coloboma syndrome and results in developmental defects of the brain, ear, eye, and kidney. *Proc Natl Acad Sci USA* 93:13870-13875.

Finnell RH, Junker WM, Wadman LK, Cabrera RM (2002) Gene expression profiling within the developing neural tube. *Neurochem Res* 27:1165-1180.

Fisher E and Scambler P (1994) Human haploinsufficiency--one for sorrow, two for joy. *Nat Genet* 7:5-7.

Fleming A, Copp AJ (1998) Embryonic folate metabolism and mouse neural tube defects. *Science* 280:2107-2109.

Footz TK, Birren B, Minoshima S, Asakawa S, Shimizu N, Riazi MA, McDermid HE (1998) The gene for death agonist *BID* maps to the region of human 22q11.2 duplicated in cat eye syndrome chromosomes and to mouse chromosome 6. *Genomics* 51:472-475.

Footz TK (1999) Mapping of the Region of Mouse Chromosome 6 Homologous to the Human Cat Eye Syndrome Critical Region. MSc Thesis, University of Alberta.

Footz TK, Brinkman-Mills P, Banting GS, Maier SA, Riazi Ma, Bridgland L, Hu S, Birren B, Minoshima S, Shimizu N, Pan H, Nguyen T, Fang F, Fu Y, Ray L, Wu H, Shaull S, Phan S, Yao Z, Chen F, Huan A, Hu P, Wang Q, Loh P, Qi S, Roe BA, McDermid HE (2001) Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: a search for candidate genes at or near the human chromosome 22 pericentromere. *Genome Res* 11:1053-1070.

Fraccaro M, Lindsten J, Ford CE, Iselius L. *et al.* (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. *Hum Genet* 56:21-51.

Fraser FC (1976) The multifactorial/threshold concept -- uses and misuses. *Teratology* 14:267-280.

Fraser FC and Nora JJ (1986) *Genetics of man.* 2nd ed. Lea & Febiger, Philadelphia, PA

Funke B, Edelmann L, McCain N, Pandita RK, Ferreira J, Merscher S, Zohouri M, Cannizzaro L, Shanske A, Morrow BE (1999) Der(22) syndrome and velo-cardio-facial syndrome/DiGeorge syndrome share a 1.5-Mb region of overlap on chromosome 22q11. *Am J Hum Genet* 64:747-758.

Giandomenico V, Simonsson M, Gronroos E, Ericsson J (2003) Coactivator-dependent acetylation stabilizes members of the SREBP family of transcription factors. *Mol Cell Biol* 23:2587-2599.

Grant PA (2001) A tale of histone modifications. *Genome Biol* 2(4):REVIEWS0003.

Greene ND and Copp AJ (1997) Inositol prevents folate-resistant neural tube defects in mouse. *Nat Med* 3:6-66.

Greene NE, Dunlevy LE, Copp AJ (2003) Homocysteine is embryotoxic but does not cause neural tube defects in mouse embryos. *Anat Embryol* (Berl) 206:185-191.

Halford S, Lindsay E, Nayudu M, Carey AH, Baldini A, Scambler PJ (1993) Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome loci at 22q11. *Hum Mol Genet* 2:191-196.

Haynes SR, Dollard C, Winston F, Beck S, Trowsdale J, Dawid IB (1992) The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucl Acids Res* 20:2603.

He B, Minges JT, Lee LW, Wilson EM (2002) The FXXLF motif mediates androgen receptor-specific interactions with coregulators. *J Biol Chem* 277: 10226-10235.

Helwig U, Imai K, Schmahl W, Thomas BE, Varnum DS, Nadeau JH, Balling R (1995) Interaction between *undulated* and *Patch* leads to an extreme form of spina bifida in double-mutant mice. *Nat Genet* 11:60-63.

Heery DM, Kalkhoven E, Hoare S, Parker MG (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387:733-738.

Henikoff S, Henikoff JG, Pietrokovski S (1999) Blocks+: A non-redundant database of protein alignment blocks dervied from multiple compilations *Bioinformatics* 15:471-479.

Herrera JE, Bergel M, Yang XJ, Nakatani Y, Bustin M (1997) The histone acetyltransferase activity of human GCN5 and PCAF is stabilized by coenzymes. *J Biol Chem* 272:27253-27258.

Hildebrand JD, Soriano P (1999) Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* 99:485-497.

Hogan B, Beddington R, Costantini F, Lacy E (1994) *Manipulating the mouse embryo: A laboratory manual* (2nd ed). Cold Spring Harbor Laboratory Press, Plainview NY.

Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, Garcia Delgado C, Gonzalez-Ramos M, Kline AD, Jabs EW (1997) Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat Genet* 15:42-46.

Huxley C, Passage E, Manson A, Putzu G, Figarella-Branger D, Pellissier JF, Fontes M (1996) Construction of a mouse model of Charcot-Marie-Tooth disease type 1A by pronuclear injection of human YAC DNA. *Hum Mol Genet* 5:563-569.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

Jacob AL, Lund J, Martinez P, Hedin L (2001) Acetylation of steroidogenic factor 1 protein regulates its transcriptional activity and recruits the coactivator GCN5. *J Biol Chem* 276:37659-37664.

Jeanmougin F, Wurtz JM, Le Douarin B, Chambon P, Losson R. (1997) The bromodomain revisited. *Trends Biochem Sci* 22:151-153.

Jhaveri MS, Wagner C, Trepel JB (2001) Impact of extracellular folate levels on global gene expression. *Mol Pharmacol* 60:1288-1295.

Johnson A, Minoshima S, Asakawa S, Shimizu N, Shizuya H, Roe BA, McDermid HE (1999) A 1.5-Mb contig within the cat eye syndrome critical region at human chromosome 22q11.2. *Genomics* 57:306-309.

Joosten PH, Hol FA, van Beersum SE, Peters H, Hamel BC, Afink GB, van Zoelen EJ, Mariman EC (1998) Altered regulation of platelet-derived growth factor receptor-

alpha gene-transcription *in vitro* by *spina bifida*-associated mutant Pax1 proteins. *Proc Natl Acad Sci USA* 95:14459-14463.

Juriloff DM and Harris MJ (2000) Mouse models for neural tube closure defects. *Hum Mol Genet* 9:993-1000.

Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen EM, Bornholdt D, Engel H, Heuer H, Aalfs CM, Ausems MG, Barone R, Herzog A, Heutink P, Homfray T, Gillessen-Kaesbach G, Konig R, Kunze J, Meinecke P, Muller D, Rizzo R, Strenge S, Superti-Furga A, Grzeschik KH. (1999) Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. *Hum Mol Genet* 8:1769-1777.

Kalter H (2000) Folic acid and human malformations: a summary and evaluation. *Reprod Toxicol* 14:463-476.

Karet FE, Finberg KE, Nelson RD, Nayir A, Mocan H, Sanjad SA, Rodriguez-Soriano J, Santos F, Cremers CW, Di Pietro A, Hoffbrand BI, Winiarski J, Bakkaloglu A, Ozen S, Dusunsel R, Goodyer P, Hulton SA, Wu DK, Skvorak AB, Morton CC, Cunningham MJ, Jha V, Lifton RP (1999) Mutations in the gene encoding B1 subunit of H+-ATPase cause renal tubular acidosis with sensorineural deafness. *Nat Genet* 21:84-90.

Kavaler J, Fu W, Duan H, Noll M, Posakony JW (1999) An essential role for the Drosophila Pax2 homolog in the differentiation of adult sensory organs. *Development* 126:2261-2272.

Kawasaki H, Schiltz L, Chiu R, Itakura K, Taira K, Nakatani Y, Yokoyama KK (2000) ATF-2 has intrinsic histone acetyltransferase activity which is modulated by phosphorylation. *Nature* 405:195-200.

Kirby ML, Waldo KL (1995) Neural crest and cardiovascular patterning. *Circ Res* 77:211-215.

Knoll JHM, Asamoah A, Pletcher BA, Wagstaff J (1995) Interstitial duplication of proximal 22q: phenotypic overlap with cat eye syndrome. *Am J Med Gen* 55:221-224.

Kohlbecker A, Lee AE, Schorle H (2002) Exencephaly in a subset of animals heterozygous for AP-2alpha mutation. *Teratology* 65:213-218.

Kohlhase J, Eischermann A, Reichenbach H, Froster U, Engel W (1998) Mutations in *SALL1* putative transcription factor gene cause Townes-Brocks syndrome. *Nat Genet* 18:81-83.

Korzus E, Torchia J, Rose DW, Xu L, Kurokawa R, McInerney EM, Mullen TM, Glass CK, Rosenfeld MG (1998) Transcription factor-specific requirements for coactivators and their acetyltransferase functions. *Science* 279:703-707.

Kramer R and Palmieri F (1992) in *Molecular Mechanisms in Bioenergetics* (Ernster, L, ed), pp. 359-384, Elsevier Science Publishers B.V., Amsterdam

Lakkis MM, Golden JA, O'Shea KS, Epstein JA (1999) Neurofibromin deficiency in mice causes exencephaly and is a modifier for Splotch neural tube defects. *Dev Biol* 212:80-92.

Lanier LM, Gates MA, Witke W, Menzies AS, Wehman AM, Macklis JD, Kwiatkowski D, Soriano P, Gertler FB (1999) Mena is required for neurulation and commissure formation. *Neuron* 22:313-325.

Lazzaro MA, Picketts DJ (2001) Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. *J Neurochem* 77:1145-56.

Lee DY, Hayes JJ, Pruss D, Wolfe AP (1993) A positive role for histone acetylation in transcription factor binding to nucleosomal DNA. *Cell* 72:73-84.

Lemon B and Tjian R (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* 14:2551-2569.

Levy NS, Chung S, Furneaux H, Levy AP (1998) Hypoxic stabilization of vascular endothelial growth factor mRNA by the RNA-binding protein HuR. *J Biol Chem* 273:6417-6423.

Liu L, McKeehan WL (2002) Sequence analysis of LRPPRC and its SEC1 domain interaction partners suggests roles in cytoskeletal organization, vesicular trafficking, nucleocytosolic shuttling, and chromosome activity. *Genomics* 79:124-136.

Lohnes D, Mark M, Mendelsohn C, Dolle P, Dierich A, Gorry P, Gansmuller A, Chambon P (1994) Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development* 120:2723-2748.

Maier SA, Poedmski L, Graham SW, McDermid HE, Locke J (2001) Characterization of the adenosine deaminase-related growth factor (ADGF) gene family in *Drosophila*. *Gene* 280:27-36.

Mari A, Amati F, Mingarelli R, Giannotti A, Sebastio G, Colloridi V, Novelli G, Dallapiccola B (1995) Analysis of the elastin gene in 60 patients with clinical diagnosis of Williams syndrome. *Hum Genet* 96:444-448.

McDermid HE, Duncan AM, Brasch KR, Holden JJ, Magenis E, Sheehy R, Burn J, Kardon N, Noel B, Schinzel A, *et al.* (1986) Characterization of the supernumerary chromosome in cat eye syndrome. *Science* 232:646-648.

McDermid HE, McTaggart KE, Riazi MA, Hudson TJ, Budarf ML, Emanuel BS, Bell CJ (1996) Long-range mapping and construction of a YAC contig within the cat eye syndrome critical region. *Genome Res* 6:1149-1159.

McInerney EM, Rose DW, Flynn SE, Westin S, Mullen TM, Krones A, Inostroza J, Torchia J, Nolte RT, Assa-Munt N, Milburn MV, Glass CK, Rosenfeld MG (1998) Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev* 12:3357-3368.

McTaggart KE, Budarf ML, Driscoll DA, Emanuel BS, Ferreira P, McDermid HE. (1998) Cat eye syndrome chromosome breakpoint clustering: identification of two intervals also associated with 22q11 deletion syndrome breakpoints. *Cytogenet Cell Genet* 81:222-228.

Mears AJ, Duncan AMV, Budarf ML, Emanuel BS, Sellinger B, Siegel-Bartelt J, Greenberg CR, McDermid HE (1994) Molecular characterization of the marker chromosome associated with cat eye syndrome. *Am J Hum Genet* 55:134-142.

Mears AJ, El-Shanti H, Murray JC, McDermid HE, Patil SR. (1995) Minute supernumerary ring chromosome 22 associated with cat eye syndrome: further delineation of the critical region. *Am J Hum Genet* 57:667-673.

Melcher M, Schmid M, Aagaard L, Selenko P, Laible G, Jenuwein T (2000) Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Mol Cell Biol* 20:3728-3741.

Mills JL, McPartlin JM, Kirke PN, Lee YJ, Conley MR, Weir DG, Scott JM (1995) Homocysteine metabolism in pregnancies complicated by neural-tube defects. *Lancet* 345:149-151.

Miner JH, Cunningham J, Sanes JR (1998) Roles for laminin in embryogenesis: exencephaly, syndactyly, and placentopathy in mice lacking the laminin alpha5 chain. *J Cell Biol* 143:1713-1723.

Misra RP and Duncan SA (2002) Gene targeting in the mouse: Advances in introduction of transgenes into the genome by homologous recombination. *Endocrine* 19:229-238.

Moore KL and Persaud TVN (1993) Before we are born: essentials of embryology and birth defects, 4th ed. WB Saunders, Philadelphia, PA.

Morris CA, Mervis CB (2000) Williams syndrome and related disorders. *Annu Rev Genomics Hum Genet* 1:461-484.

Morrison K, Papapetrou C, Hol FA, Mariman EC, Lynch SA, Burn J, Edwards YH (1998) Susceptibility to spina bifida; an association study of five candidate genes. *Ann Hum Genet* 62(Pt 5):379-396.

MRC Vitamin Study Research Group (1991) Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 338:131-137.

Nakai K and Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* 24:34-36.

Nakatsu T, Uwabe C, Shiota K (2000) Neural tube closure in humans initiates at multiple sites: evidence from human embryos and implications for the pathogenesis of neural tube defects. *Anat Embryol* (Berl) 201:455-466.

Nesslinger NJ, Gorski JL, Kurczynski TW, Shapira SK, Siegel-Bartelt J, Dumanski JP, Cullen RF Jr, French BN, McDermid HE (1993) Clinical, cytogenetic, and molecular characterization of seven patients with deletions of chromosome 22q13.3. *Am J Hum Genet* 54:464-472.

Netzer C, Rieger L, Brero A, Zhang CD, Hinzke M, Kohlhase J, Bohlander SK (2001) *SALL1*, the gene mutated in Townes-Brocks syndrome, encodesa transcriptional repressor which interacts with TRF1/PIN2 and localizes to pericentromeric heterochromatin. *Hum Mol Genet* 10(26):3017-3024.

Nowak SJ, Corces VG (2000) Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes Dev* 14:3003-3013.

Nussbaum RL, McInnes RR, Willard, HF (2002) *Genetics in Medicine* 6th ed., WB Saunders, Philadelphia, PA.

Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y (1996) The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87:953-959.

Ohura T, Kobayashi K, Tazawa Y, Nishi I, Abukawa D, Sakamoto O, Iinuma K, Saheki T (2001) Neonatal presentation of adult-onset type II citrullinemia. *Hum Genet* 108:87-90.

Okabe I, Bailey LC, Attree O, Srinivasan S, Perkel JM, Laurent BC, Carlson M, Nelson DL, Nussbaum RL (1992) Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast S. cerevisiae. *Nucleic Acids Res* 20:4649-4655

O'Rahilly R, Muller F (2002) The two sites of fusion of the neural folds and the two neuropores in the human embryo. *Teratology* 65:162-170.

Pagon RA, Graham JM, Zonana J, Young SL (1981) Coloboma, congenital heart disease and choanal atresia with multiples anomalies: CHARGE association. *J Pediatr* 99:223-227.

Palmieri F (1994) Mitochondrial carrier proteins. *FEBS Lett* 346:48-54.

Palmieri L, Runswick MJ, Fiermonte G, Walker JE, Palmieri F (2000) Yeast mitochondrial carriers: bacterial expression, biochemical identification and metabolic significance. *J Bioenerg Biomembr*. 32:67-77.

Pani L, Horal M, Loeken MR (2002) Rescue of neural tube defects in Pax-3-deficient embryos by p53 loss of function: implications for Pax-3- dependent development and tumorigenesis. *Genes Dev* 16:676-680.

Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, Masuno M, Tommerup N, van Ommen GJ, Goodman RH, Peters DJ, *et al.* (1995) Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* 376:348-351.

Pham AD and Sauer F (2000) Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in Drosophila. *Science* 289:2357-2360.

Phelan MC, Rogers RC, Saul RA, Stapleton GA, Sweet K, McDermid H, Shaw SR, Claytor J, Willis J, Kelly DP (2001) 22q13 deletion syndrome. *Am J Med Genet* 101:91-99.

Phelan SA, Ito M, Loeken MR (1997) Neural tube defects in embryos of diabetic mice: role of the Pax-3 gene and apoptosis. *Diabetes* 46:1189-1197.

Piedrahita JA, Oetama B, Bennett GD, van Waes J, Kamen BA, Richardson J, Lacey SW, Anderson RG, Finnell RH (1999) Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development. *Nat Genet* 23:228-232.

Powell CM and RC Michaelis (1999) Townes-Brocks syndrome. *J Med Genet* 36:89-93.

Qutob MS, Bhattacharjee RN, Pollari E, Yee SP, Torchia J (2002) Microtubule-dependent subcellular redistribution of the transcriptional coactivator p/CIP. *Mol Cell Biol* 22:6611-6626.

Reeves R, Nissen MS (1990) The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J Biol Chem* 265:8573-8582.

Reiss JA, Weleber RG, Brown MG, Bangs CD, Lovrien EW, Magenis RE (1985) Tandem duplication of proximal 22q: A cause of cat-eye syndrome. *Am J Med Genet* 20:165-171.

Riazi MA, Brinkman-Mills P, Nguyen T, Pan H, Phan S, Ying F, Roe BA, Tochigi J, Shimizu Y, Minoshima S, Shimizu N, Buchwald M, McDermid H. (2000) The

human homolog of insect-derived growth factor, *CECR1*, is a candidate gene for features of cat eye syndrome. *Genomics* 64:277-285.

Rosenberg MJ, Agarwala R, Bouffard G, Davis J, Fiermonte G, Hilliard MS, Koch T, Kalikin LM, Makalowska I, Morton DH, Petty EM, Weber JL, Palmieri F, Kelley RI, Schaffer AA, Biesecker LG (2002) Mutant deoxynucleotide carrier is associated with congenital microcephaly. *Nat Genet* 32:175-179.

Rosenquist TH, Ratashak SA, Selhub J (1996) Homocysteine induces congenital defects of the heart and neural tube: effect of folic acid. *Proc Natl Acad Sci USA* 93:15227-15232.

Sabapathy K, Jochum W, Hochedlinger K, Chang L, Karin M, Wagner EF (1999) Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. *Mech Dev* 89:115-124.

Sambrook J and Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edition (Cold Spring Harbor Press, Cold Spring Harbor, NY)

Sanyanusin P, Schimmenti LA, McNoe LA, Ward TA, Pierpont MEM, Sullivan MJ, Dobyns WB, Eccles MR (1995) Mutation of the *PAX2* gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nature Genet.* 9:358-364.

Schimmenti LA, Pierpont ME, Carpenter BLM, Kashtan CE, Johnson MR, Dobyns WB (1995) Autosomal dominant optic nerve colobomas, vesicoureteral reflux, and renal anomalies. *Am. J. Med. Genet.* 59:204-208.

Schinzel A, Schmid W, Fraccaro M, Tiepolo L, Zuffardi O, Opitz JM, Lindsten J, Zetterqvist P, Enell H, Baccichetti C, Tenconi R, Pagon RA (1981a) The "cat

eye syndrome": dicentric small marker chromosome probably derived from a no.22 (tetrasomy 22pter-q11) associated with a characteristic phenotype. *Hum Genet* 57:148-158.

Schinzel A, Schimd W, Auf der Maur P, Moser H, Degenhardt KH, Geisler M, Grubisic A (1981b) Incomplete trisomy 22. I. Familial 11/22 translocation with 3:1 meiotic disjunction. Delineation of a common clinical picture and report of nine new cases from six families. *Hum Genet* 56:249-262.

Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell PJ (1996) Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* 381:235-238.

Schultz DC, Friedman JR, Rauscher FJ 3rd (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev* 15:428-443.

Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, Driscoll DA, McDonald-McGinn DM, Zackai EH, Budarf ML, Emanuel BS (1999a) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* 9:489-501.

Shaikh TH, Budarf ML, Celle L, Zackai EH, Emanuel BS (1999b) Clustered 11q23 and 22q11 breakpoints and 3:1 meiotic malsegregation in multiple unrelated t(11;22) families. *Am J Hum Genet* 65:1595-1607.

Shamay M, Barak O, Shaul Y (2002) HBXAP, a novel PHD-finger protein, possesses transcription repression activity. *Genomics* 79:523-529

Shields DC, Kirke PN, Mills JL, Ramsbottom D, Molloy AM, Burke H, Weir DG, Scott JM, Whitehead AS (1999) The "thermolabile" variant of methylenetetrahydrofolate reductase and neural tube defects: An evaluation of genetic risk and the relative importance of the genotypes of the embryo and the mother. *Am J Hum Genet* 64:1045-1055.

Shprintzen RJ, Goldberg RB, Lewin ML, Sidoti EJ, Berkman MD, Argamaso RV, Young D (1978) A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. *Cleft Palate J* 15:56-62.

Smith DJ, Stevens ME, Sudanagunta SP, Bronson RT, Makhinson M, Watabe AM, O'Dell TJ, Fung J, Weier HU, Cheng JF, Rubin EM (1997) Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects associated with Down syndrome. *Nat Genet* 16:28-36.

Smith AN, Skaug J, Choate KA, Nayir A, Bakkaloglu A, Ozen S, Hulton SA, Sanjad SA, Al-Sabban EA, Lifton RP, Scherer SW, Karet FE (2000) Mutations in ATP6N1B, encoding a new kidney vacuolar proton pump 116-kD subunit, cause recessive distal renal tubular acidosis with preserved hearing. *Nat Genet* 26:71-75.

Southern EM. (1974) An improved method for transferring nucleotides from electrophoresis strips to thin layers of ion-exchange cellulose. *Anal Biochem* 62:317-318.

Strahl BD, Ohba R, Cook RG, Allis CD (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc Natl Acad Sci USA* 96:14967-14972.

Steegers-Theunissen RP, Boers GH, Trijbels FJ, Finkelstein JD, Blom HJ, Thomas CM, Borm GF, Wouters MG, Eskes TK. (1994) Maternal hyperhomocysteinemia: a risk factor for neural-tube defects? *Metabolism* 43:1475-1480.

Stegmann K, Boecker J, Richter B, Capra V, Finnell RH, Ngo ET, Strehl E, Ermert A, Koch MC (2001) A screen for mutations in human homologues of mice exencephaly genes Tfap2alpha and Msx2 in patients with neural tube defects. *Teratology* 63:167-175.

Stephenson DA, Mercola M, Anderson E, Wang CY, Stiles CD, Bowen-Pope DF, Chapman VM (1991) Platelet-derived growth factor receptor alpha-subunit gene (Pdgfra) is deleted in the mouse patch (Ph) mutation. *Proc Natl Acad Sci USA* 88:6-10.

Strachan T and Read AP (1996) *Human Molecular Genetics*. BIOS Scientific Publishers Ltd., Oxford, UK

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 2000 403:41-45.

Suter U, Moskow JJ, Welcher AA, Snipes GJ, Kosaras B, Sidman RL, Buchberg AM, Shooter EM (1992a) A leucine-to-proline mutation in the putative first transmembrane domain of the 22-kDa peripheral myelin protein in the trembler-J mouse. *Proc Nat Acad Sci USA* 89:4382-4386.

Suter U, Welcher AA, Ozcelik T, Snipes GJ, Kosaras B, Francke U, Billings-Gagliardi S, Sidman RL, Shooter EM (1992b) Trembler mouse carries a point mutation in a myelin gene. *Nature* 356:241-244.

Swan DG, Stern R, McKellar S, Phillips K, Oura CA, Karagenc TI, Stadler L, Shiels BR (2001) Characterisation of a cluster of genes encoding Theileria annulata AT hook DNA-binding proteins and evidence for localisation to the host cell nucleus. *J Cell Sci* 114:2747-2754.

Tanaka Y, Naruse I, Maekawa T, Masuya H, Shiroishi T, Ishii S (1997) Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein-Taybi syndrome. *Proc Natl Acad Sci U S A* 94:10215-10220.

Tanese N, Pugh BF, Tjian R (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes & Dev* 5:2212-2224.

Tate P, Lee M, Tweedie S, Skarnes WC, Bickmore WA (1998) Capturing novel mouse genes encoding chromosomal and other nuclear proteins. *J Cell Sci* 111:2575-2585.

Tellier AL, Amiel J, Delezoide AL, Audollent S, Auge J, Esnault D, Echa-Razavi F, Munnich A, Lyonnet S, Vekemans M, Attie-Bitach T (2000) Expression of the *PAX2* gene in human embryos and exclusion in the CHARGE syndrome. *Am J Med Genet* 93:85-88.

Thomson S, Mahadevan LC, Clayton AL (1999) MAP kinase-mediated signalling to nucleosomes and immediate-early gene induction. *Semin Cell Dev Biol* 10:205-214.

Torres M, Gomez-Pardo E, Dressler GR, Gruss P (1995) Pax-2 controls multiple steps of urogenital development. *Development* 121:4057-4065

Torres M, Gomez-Pardo E, Gruss P (1996) Pax2 contributes to inner ear patterning and optic nerve trajectory. *Development* 122:3381-3391

Tran P, Hiou-Tim F, Frosst P, Lussier-Cacan S, Bagley P, Selhub J, Bottiglieri T, Rozen R (2002) The curly-tail (*ct*) mouse, an animal model of neural tube defects, displays altered homocysteine metabolism without folate responsiveness or a defect in *Mthfr*. *Mol Genet Metab* 76:297-304.

Trembath D, Sherbondy AL, Vandyke DC, Shaw GM, Todoroff K, Lammer EJ, Finnell RH, Marker S, Lerner G, Murray JC (1999) Analysis of select folate pathway genes, PAX3, and human T in a Midwestern neural tube defect population. *Teratology* 59:331-341.

Tronche F, Casanova E, Turiault M, Sahly I, Kellendonk C (2002) When reverse genetics meets physiology: the use of site-specific recombinases in mice. *FEBS Lett* 529:116-121.

Tse C, Sera T, Wolffe AP, Hansen JC (1998) Disruption of higher order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* 18:4629-4638.

Tsukiyama T, Palmer J, Landel CC, Shiloach J, Wu C (1999) Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. *Genes Dev* 13:686-697

van Aerts LA, Blom HJ, Deabreu RA, Trijbels FJ, Eskes TK, Peereboom-Stegeman JH, Noordhoek J. (1996) Prevention of neural tube defects by and toxicity of L-homocysteine in cultured postimplantation rat embryos. *Teratology* 50:348-360.

van der Put NM, Steegers-Theunissen RP, Frosst P, Trijbels FJ, Eskes TK, van den Heuvel LP, Mariman EC, den Heyer M, Rozen R, Blom HJ (1995) Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. *Lancet* 346:1070-1071.

van der Put NM, van den Heuvel LP, Steegers-Theunissen RP, Trijbels FJ, Eskes TK, Mariman EC, den Heyer M, Blom HJ (1996) Decreased methylene tetrahydrofolate reductase activity due to the 677C-->T mutation in families with spina bifida offspring. *J Mol Med* 74:691-694.

van Hove JL, McConkie-Rosell A, Chen YT, Iafolla AK, Lanman JT Jr, Hennessy MD, Kahler SG (1992) Unbalanced translocation 46,XY,-15, +der(22) t(15;22) (q13;q11)pat: case report and review of the literature. *Am J Med Genet* 44:24-30.

van Straaten HW, Hekking JW, Consten C, Copp AJ (1993) Intrinsic and extrinsic factors in the mechanism of neurulation: effect of curvature of the body axis on closure of the posterior neuropore. *Development* 117:1163-1172.

Vortkamp A, Gessler M, Grzeschik KH (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* 352:539-540.

Webb P, Anderson CM, Valentine C, Nguyen P, Marimuthu A, West BL, Baxter JD, Kushner PJ (2000) The nuclear receptor corepressor (N-CoR) contains three

isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs). *Mol Endocrinol* 14:1976-1985.

Williams RS, Wagner PD (2000) Transgenic animals in integrative biology: approaches and interpretations of outcome. *J Appl Physiol* 88:1119-1126.

Wilson A, Platt R, Wu Q, Leclerc D, Christensen B, Yang H, Gravel RA, Rozen R (1999) A common variant in methionine synthase reductase combined with low cobalamin (vitamin B12) increases risk for spina bifida. *Mol Genet Metab* 67:317-323.

Wilson V, Manson L, Skarnes WC, Beddington RS (1995) The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* 121:877-886.

Wu M, Chen DF, Sasaoka T, Tonegawa S (1996) Neural tube defects and abnormal brain development in F52-deficient mice. *Proc Natl Acad Sci USA* 93:2110-2115.

Xu W, Baribault H, Adamson ED (1998) Vinculin knockout results in heart and brain defects during embryonic development. *Development* 125:327-337.

Xue Y, Wong J, Moreno GT, Young MK, Cote J, Wang W (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* 2:851-861.

Yang XJ, Ogryzko VV, Nishikawa J, Howard BH, Nakatani Y (1996) A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* 382:319-324.

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R (1998) Gene dosage-dependent embryonic

development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 93:361-372.

Yao Z, Painter SL, Fanslow WC, Ulrich D, Macduff BM, Spriggs MK, Armitage RJ (1995) Human IL-17: a novel cytokine from T cells. *J Immunol* 155:5483-5486.

Yao Z, Spriggs MK, Derry JM, Strockbine L, Park LS, VandenBos T, Zappone JD, Painter SL, Armitage RJ (1997) Molecular characterization of the human interleukin (IL)-17 receptor. *Cytokine* 9:794-800.

Yeap BB, Voon DC, Vivian JP, McCulloch RK, Thomson AM, Giles KM, Czyzyk-Krzeska MF, Furneaux H, Wilce MC, Wilce JA, Leedman PJ (2002) Novel binding of HuR and poly(C)-binding protein to a conserved UC-rich motif within the 3'-untranslated region of the androgen receptor messenger RNA. *J Biol Chem* 277:27183-27192.

Yin Z, Haynie J, Yang X, Han B, Kiatchoosakun S, Restivo J, Yuan S, Prabhakar NR, Herrup K, Conlon RA, Hoit BD, Watanabe M, Yang YC (2002) The essential role of Cited2, a negative regulator for HIF-1alpha, in heart development and neurulation. *Proc Natl Acad Sci USA*. 99:10488-93.

Zackai EH and Emanuel BS (1980) Site-specific reciprocal translocation, t(11;22) (q23;q11), in several unrelated families with 3:1 meiotic disjunction. *Am J Med Genet* 7:507-521.

Zhao Q, Behringer RR, de Crombrugghe B (1996) Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the *Cart1* homeobox gene. *Nat Genet* 13:275-283.

Zheng B, Sage M, Sheppeard EA, Jurecic V, Bradley A (2000) Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol Cell Biol* 20:648-655.

Zurovec M, Dolezal T, Gazi M, Pavlova E, Bryant PJ (2002) Adenosine deaminase-related growth factors stimulate cell proliferation in *Drosophila* by depleting extracellular adenosine. *Proc Natl Acad Sci USA* 99:4403-4408.

APPENDIX

Addresses for internet sites used in this study:

Repeatmasker Web Server
http://ftp.genome.washington.edu/cgi-bin/RepeatMasker

BLAST
http://www.ncbi.nlm.nih.gov/BLAST/

Online Mendelian Inheritance in Man
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

GENSCAN gene prediction
http://genes.mit.edu/GENSCAN.html

MOTIF
http://motif.genome.ad.jp/

Expasy protein manipulation suite
www.expasy.org

Flybase
www.flybase.org

Table A1. Primers for human genes discussed in this study.

Primer	Location	Sequence (5' to 3')
CECR2 F33 (Bam HI)	*CECR2* exon 1	cgc agc gga tcc atg tgc cca gag gag ggc ggc
CECR2 F21	*CECR2* exon 1	gcc atc gcg cac ttc tgc tc
CECR2 F22	*CECR2* exon 1	ggc ggc agc agc ggg agg agc
CECR2 F23	*CECR2* exon 1	gcc cct ccg ccc cta gcc cca tct gt
CECR2 F24	*CECR2* exon 1	ggg ggt tgt tgt tgt ggc gcg ggc ag
CECR2 R21	*CECR2* exon 1	tct cga agt cgg gca ggc gga ac
CECR2 R22	*CECR2* exon 1	gcg gtg cga aag agc gag cag aag
CECR2 R24	*CECR2* exon 1	cgg ccg ccc cca gtc ccc tac aa
CECR2 R17	*CECR2* exon 2	agg cag gca atc agg tca ctg
CECR2 F14	*CECR2* exon 3	ccg gct gga tgc aga cga tgt
CECR2 R14	*CECR2* exon 3	ccc tca gag ggt tgg gct tcc
CECR2 R15	*CECR2* exon 3	acg agc tcc cag cgg tag ttg
CECR2 R19	*CECR2* exon 4	cac acg gag act gtc tgc atc
CECR2 R20	*CECR2* exon 4	cca gaa ttg tct tca ccc aat ggc tc
CECR2 F16	*CECR2* exon 5	ctg cag gag gag att ctg ttg
CECR2 R23	*CECR2* exon 6	gga tgc caa gga att ttc ttc ctg
CECR2 F29	*CECR2* exon 7	ggt act tgg tgg ctc ctg tgc
CECR2 F30	*CECR2* exon 7	gga ctt cct gcc tga gat ctg
CECR2 F1	*CECR2* exon 10	atc gag cga aga gga gaa agc
CECR2 F2	*CECR2* exon 10	tgg ctg ctg gct caa gga aag
CECR2 R7	*CECR2* exon 10	ctt tcc ttg agc cag cag cca
CECR2 R11	*CECR2* exon 10	gct ttc tcc tct tcg ctc gat
CECR2 R26	*CECR2* exon 13'	ggc att tta ttc ttg att aca g
CECR2 R27	*CECR2* exon 13'	gta tct agc aca aac tcc ttc cc
CECR2 F6	*CECR2* exon 14	gat cag agc agc agc tcc aca
CECR2 R1'	*CECR2* exon 14	tgt gga gct gct gct ctg atc
CECR2 F4	*CECR2* exon 15	gat cct gcc acc ttg tat ggc
CECR2 R2'	*CECR2* exon 15	ggt gaa agg ctg acg ctg ctg
CECR2 F32 (Bam HI)	*CECR2* exon 16	gcc cac acc ctg gat cct tgc agc tt
CECR2 R32 (Bam HI)	*CECR2* exon 16	aag ctg caa gga tcc agg gtg tgg gc
CECR2 F5.1	*CECR2* exon 16	cag ttc cag cca gga ttc att

Primer	Location	Sequence (5' to 3')
CECR2 F3	*CECR2* exon 17	gac tgc acc agg cag agc tca
CECR2 F5	*CECR2* exon 17	gca ctg tga gcc agt ttc ccc
CECR2 F7	*CECR2* exon 17	gaa ttc agc tgc cca tta cca
CECR2 F8	*CECR2* exon 17	ggc cag aga gtc cca aag aat
CECR2 R3	*CECR2* exon 17	tga gct ctg cct ggt gca gtc
CECR2 R10	*CECR2* exon 17	ggg gaa act ggc tca cag tgc
CECR2 R12	*CECR2* exon 17	ccg ctt ggt agc tgc gtt atg
CECR2 F9	*CECR2* exon 18	cag tcg cag gcc tcg ttc cca
CECR2 R5'	*CECR2* exon 18	tgg gaa cga ggc ctg cga ctg
CECR2 R33 (Xho I)	*CECR2* exon 18	cta gct ctg ctc gag ggg aag tgt tgg
CECR2 F10	*CECR2* exon 19	atg aac gtc gag tta atg atg
CECR2 F11	*CECR2* exon 19	tga tga gca gct aaa ctc agg
CECR2 R6	*CECR2* exon 19	cca gca tga gag agc agc ctg
CECR2 R30 (3' UTR)	*CECR2* exon 19"	gcc cct ccc aca gct gaa ttt ata ag
CECR2 F12 (3" UTR)	*CECR2* exon 19"	cag agg cct ctc act tct cct
CECR2 F13 (3' UTR)	*CECR2* exon 19"	cag gag cac tga cac tgt ggg
CECR2 F14 (3' UTR)	*CECR2* exon 19"	gga agt agt gtg taa ggt atc c
CECR2 F15 (3' UTR)	*CECR2* exon 19"	cac cct gtg ggt cac tgt cac
CECR2 F16 (3' UTR)	*CECR2* exon 19"	ccc ggt cag tcc cca gca gcc ttg
CECR2 F17 (3' UTR)	*CECR2* exon 19"	tct gga gcc tct gct cct gct gca
CECR2 F25 (3' UTR)	*CECR2* exon 19"	tgc ccc tca gtt cag tca gac ctc ag
CECR2 F26 (3' UTR)	*CECR2* exon 19"	cca tcc cag gcc cac aaa atc cca gt
CECR2 F27 (3' UTR)	*CECR2* exon 19"	caa tcc ggg ggc tcc tgg ggt gtg
CECR2 F28 (3' UTR)	*CECR2* exon 19"	tgt gat tgc cat ttg ctg cgt gag
CECR2 R22 (3' UTR)	*CECR2* exon 19"	gct ccc acc cca ccc acc cct gtc
CECR2 R26 (3' UTR)	*CECR2* exon 19"	gct ttt gca gca gga gca gag gc
CECR2 R27 (3' UTR)	*CECR2* exon 19"	ggc ccc agg agc tac gca caa g
CECR2 R28 (3' UTR)	*CECR2* exon 19"	ttc tca cgc agc aaa tgg caa tca c
CECR2 R29 (3' UTR)	*CECR2* exon 19"	cca cac ccc agg agc ccc cgg att
GCN5 F1	*GCN5*	ctg tca agc ttg gag gag gag atc tat ggg gca
GCN5 F2	*GCN5*	gcg gct ccg tgt gat ggg tga c

Primer	Location	Sequence (5' to 3')
GCN5 F3	*GCN5*	agc ccg acg agt acg cca tc
GCN5 R1	*GCN5*	tcc ctc ctc gag ctt gaa gta gaa gaa ctt
GCN5 R2	*GCN5*	cgg ggt cct tca gct cct tc
GCN5 R3	*GCN5*	cgt cgg cgt agg tga gga agt a
77H2-1 (GSP2)	*SLC25A18* exon 1	cac tgc cta ctc agt ctc ttc t
MTP F1	*SLC25A18* exon 1	gaa ctg agt agg cag tga ga
MTP R1	*SLC25A18* exon 2	ctt tcg gta aga acc tct gc
HS7	*SLC25A18* exon 4	cct cca ttg atg agt ttg gc
HS3	*SLC25A18* exon 8	tgg aag ccg aac cag ttg tgt
NX7	*SLC25A18* exon 11	tgc cag gaa act ctg gat tca
NX3	*SLC25A18* exon 11	ctc aaa cta ggt gct gac tct

Table A2. Primers for mouse genes discussed in this study.

Primer	Location	Sequence (5'-3')
Sry FWD	Mus musculus Sry	gag agc atg gag ggc cat
Sry REV	*Mus musculus Sry*	cca ctc ctc tgt gac act
Cecr2 7i.F1	*Cecr2* Intron 7	ggc cca tgc tgt tcc ttc ctg ata g
Cecr2 7i.R1	*Cecr2* Intron 7	aag gct gac att gtg agg gcg aaa
Cecr2 F1	*Cecr2* exon 2	ttg cct gct tgc ttc aag gct gct
Cecr2 R9	*Cecr2* exon 2	gtg ata tct ctt cgc tga tag cag
Cecr2 R10	*Cecr2* exon 2	ata aac tcc acg tca tct ctg tga
Cecr2 F2	*Cecr2* exon 3	cct gca tcg cct ctg tga tta ccg
Cecr2 R8	*Cecr2* exon 3	tct act cga gta cgc aga gga agg
Cecr2 R1	*Cecr2* exon 7	tcc gct ctc gga agc tct cag tga
Cecr2 R2	*Cecr2* exon 7	aag tac ctt ggc ctg gcc ctc gg
Cecr2 F3	*Cecr2* exon 9	gcg cag gga gga gga gga gga gcg
Cecr2 F4	*Cecr2* exon 9	gcg aca gct cct tct tgc cgt gc
Cecr2 F5	*Cecr2* exon 10	caa ggc aag gag cta ccc cca gaa
Cecr2 F6	*Cecr2* exon 14	gag cgg tgt ttc cat cgg gcc at
Cecr2 R6	*Cecr2* exon 14'	gcg aga aaa gcc ttg gcc atg tga
Cecr2 F7	*Cecr2* exon 16	tcc acc agc tcc att cca ggc agg
Cecr2 R5	*Cecr2* exon 16	cct gcc tgg aat gga gct ggt gga
Cecr2 R4	*Cecr2* exon 16'	ctg ggt ctt cct cgg gct cgt gtg
Cecr2 F8	*Cecr2* exon 17	gag gct cat ttc aag aag tac
Cecr2 R3	*Cecr2* exon 17	cgc ttc cgg tac cag tct gct cca
Cecr2 R13	*Cecr2* exon 17	ctg ggt gag gta gct gtt tga ggt
Cecr2 R12	*Cecr2* exon 19"	cct cct cac acc tat act cta

Table A3. Primers for vectors used in this study.

Primer Name	Vector	Sequence (5' to 3')
T7	Various	taa tac gac tca cta tag gg
T3	Various	taa ccc tca cta aag gga
SP6	various	att tag gtg aca cta tag
OPIE2 FWD	pMIB A-C	cgc aac gat ctg gta aac ac
OPIE2 REV	pMIB A-C	gac aat aca aac taa gat tta gtç ag
pMIB polylinker fwd	pMIB polylinker modification	agc ttg cta gtc gga tcc cat atg cag tgt g
pMIB polylinker rev	pMIB polylinker modification	aat tca cac tgc ata tgg gat ccg act agc a
pGT1 F1	pGT1 genetrap vector	tcc cgt ggt ctc gcc ctc ttg tcc t
pGT1 R1	pGT1 genetrap vector	gct gcg ggg atg gtg gag gaa ac
pGT1 R4	pGT1 genetrap vector	acg cca tac agt cct ctt cac atc

Table A4. Primers for *Drosophila* genes discussed in this study.

Primer Name	Location	Sequence (5' to 3')
Su(var) 3-7 fwd	Su(var) 3-7 exon 1	ccg ctg aaa agg ggt gag ttg ac
Su(var) 3-7 rev	Su(var) 3-7 exon 1	cta gag gac tta tga gca aaa ct
Dm CECR2 F4	CG10115 intron 1	ctg gga aat ggg cgt ctc taa agt ta
Dm CECR2 R4	CG10115 intron 1	ggc gca caa tta act cgg gca cta g

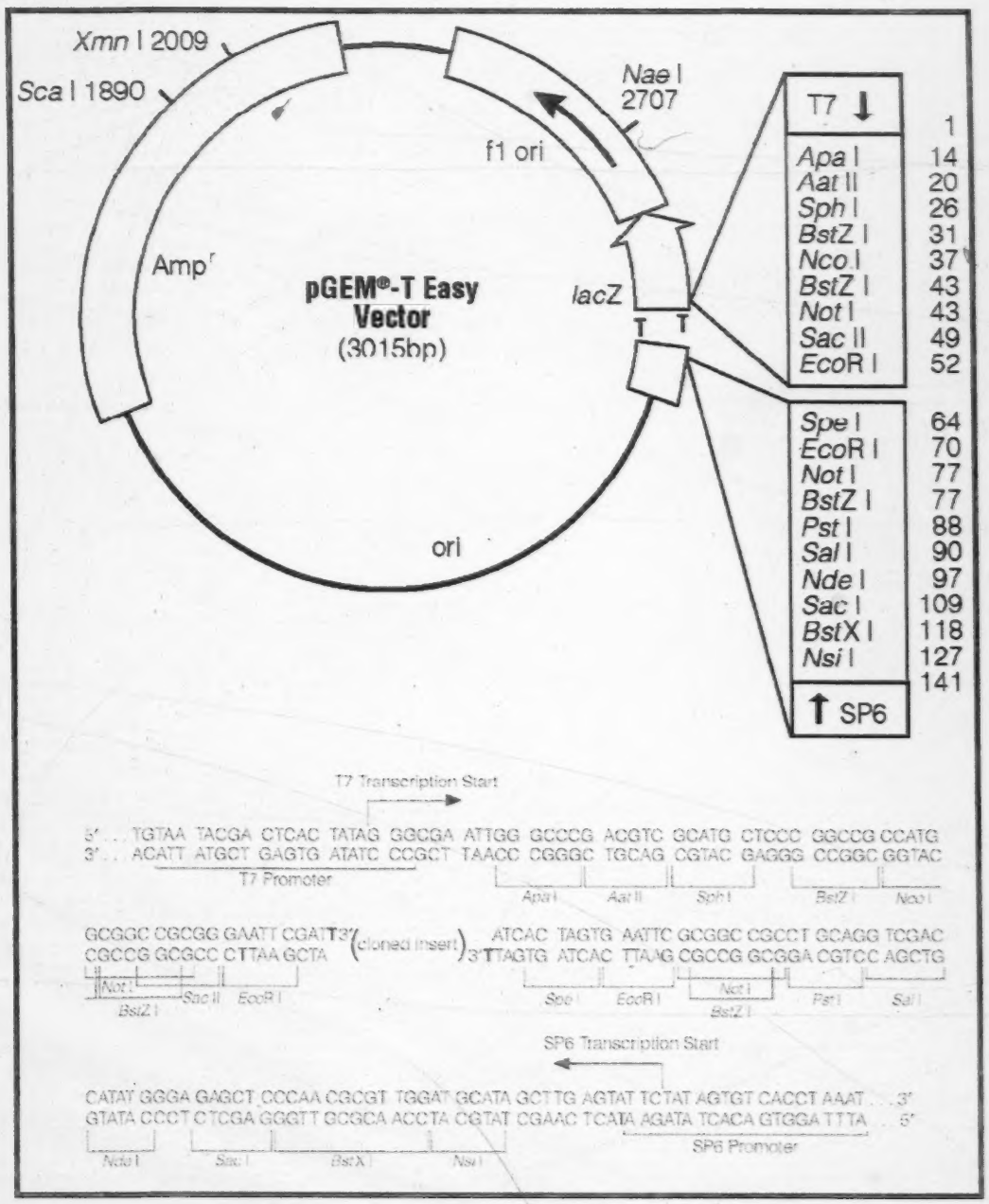

Xmn I 2009
Sca I 1890
Nae I 2707
f1 ori
Amp^r
pGEM®-T Easy Vector (3015bp)
lacZ
T T
ori
T7
Apa I 14
Aat II 20
Sph I 26
BstZ I 31
Nco I 37
BstZ I 43
Not I 43
Sac II 49
EcoR I 52
Spe I 64
EcoR I 70
Not I 77
BstZ I 77
Pst I 88
Sal I 90
Nde I 97
Sac I 109
BstX I 118
Nsi I 127
141
SP6
T7 Transcription Start
5' . . TGTAA TACGA CTCAC TATAG GGCGA ATTGG GCCCG ACGTC GCATG CTCCC GGCCG CCATG
3' . . ACATT ATGCT GAGTG ATATC CCGCT TAACC CGGGC TGCAG CGTAC GAGGG CCGGC GGTAC
T7 Promoter
Apa I Aat II Sph I BstZ I Nco I
GCGGC CGCGG GAATT CGATT 3'
CGCCG GCGCC CTTAA GCTA
(cloned insert)
ATCAC TAGTG AATTC GCGGC CGCCT GCAGG TCGAC
3' TTAGTG ATCAC TTAAG CGCCG GCGGA CGTCC AGCTG
Not I Sac II EcoR I
BstZ I
Spe I EcoR I Not I
BstZ I
Pst I Sal I
SP6 Transcription Start
CATAT GGGA GAGCT CCCAA CGCGT TGGAT GCATA GCTTG AGTAT TCTAT AGTGT CACCT AAAT . . 3'
GTATA CCCT CTCGA GGGTT GCGCA ACCTA CGTAT CGAAC TCATA AGATA TCACA GTGGA TTTA . . 5'
SP6 Promoter
Nde I Sac I BstX I Nsi I

Figure A1. Map and polylinker region of the T/A cloning vector pGEM-Teasy used for the cloning of PCR products (Promega).

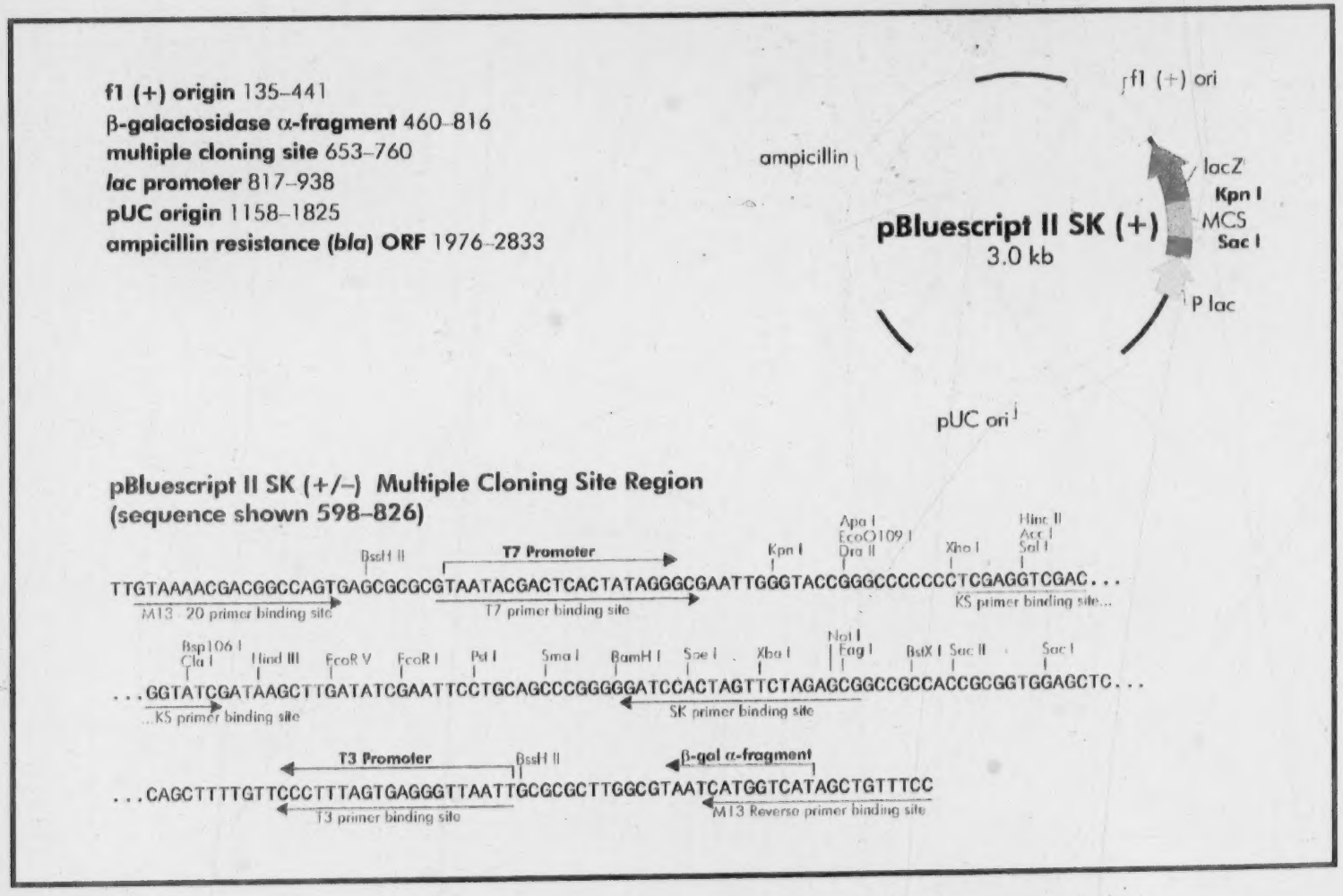


Figure A2. Map and multiple cloning site of the multipurpose cloning vector pBS-II-SK$^+$ (Stratagene).

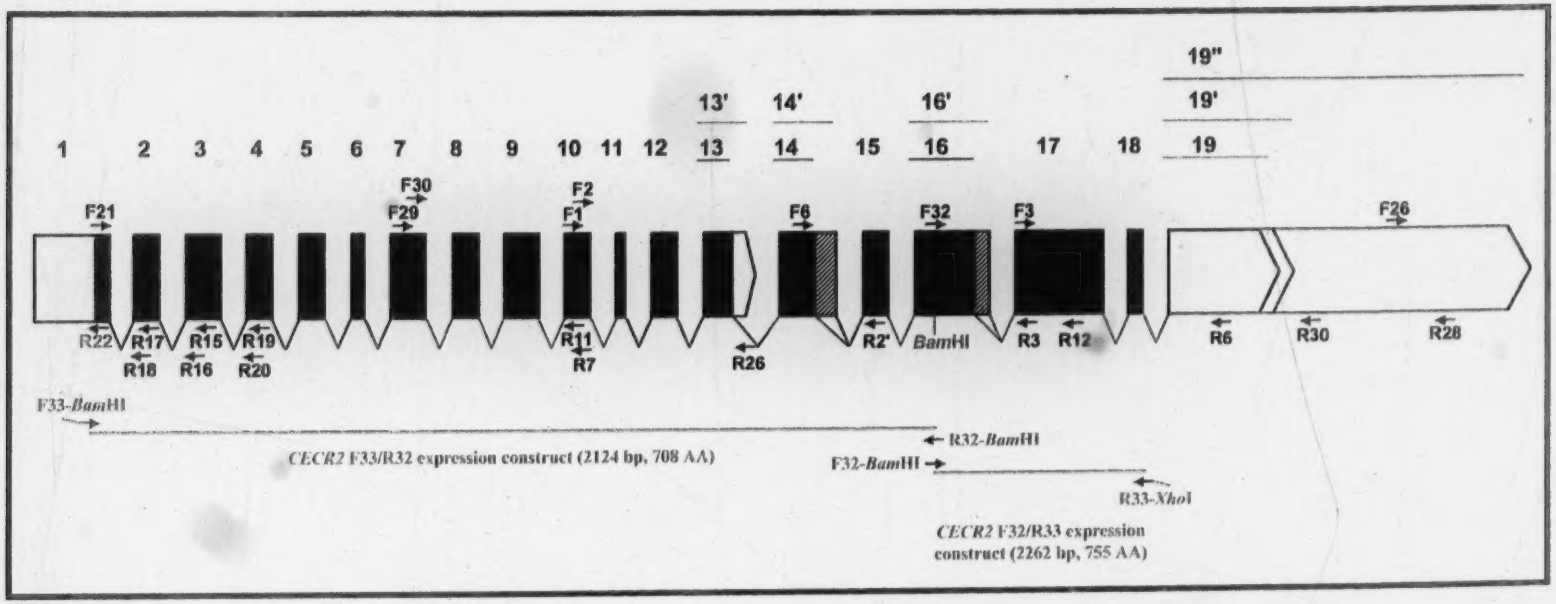


Figure A3. Schematic of *CECR2* (not to scale) showing the position of the expression constructs generated from *CECR2*. The outer primers (F33, R33) utilized tails which contained restriction sites to facilitate cloning of the products. The inner primers (F32, R32) spanned a native *Bam*HI site within *CECR2*.

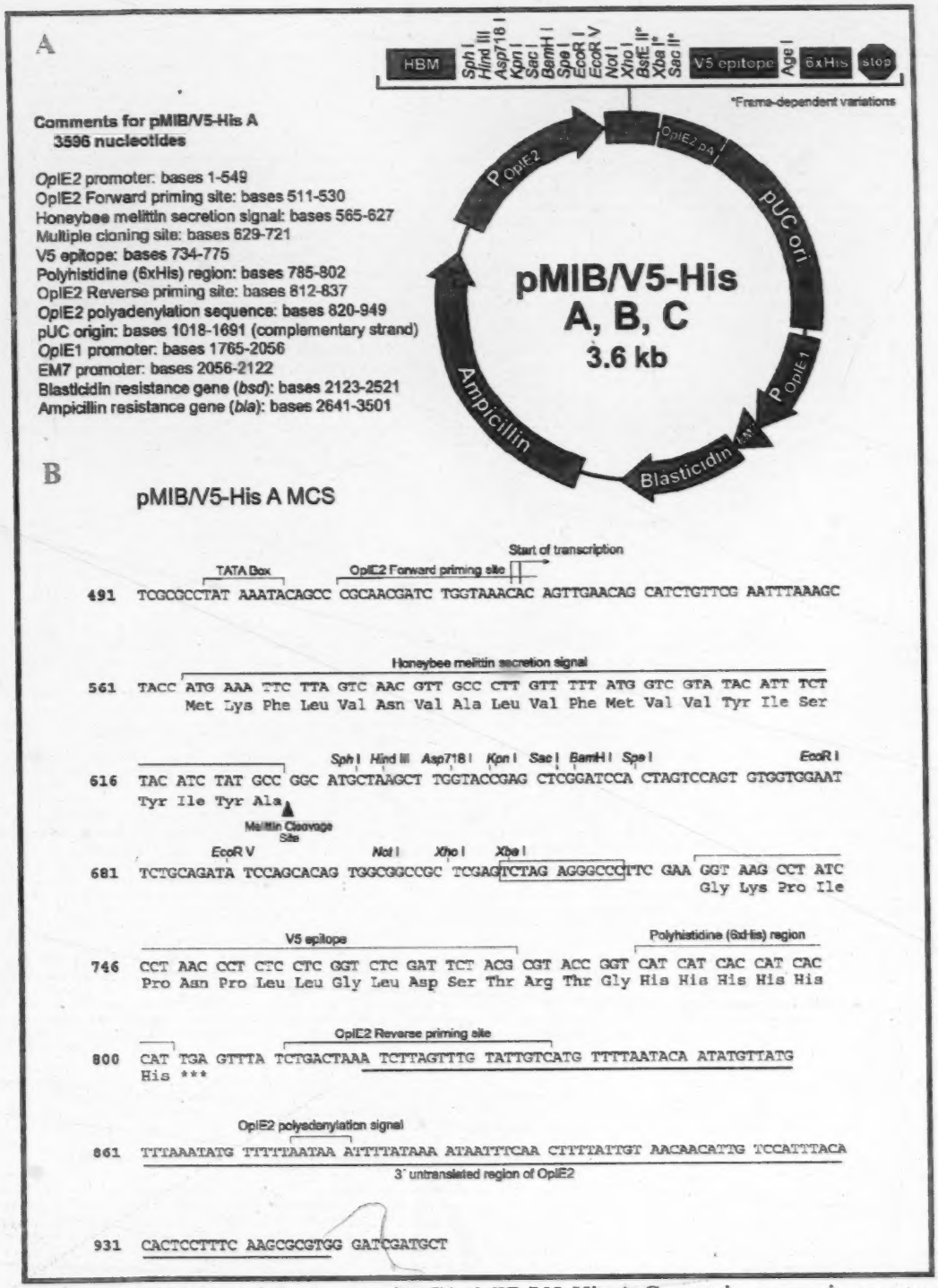

A
HBM
Sph I
Hind III
Asp718 I
Kpn I
Sac I
BamH I
Spe I
EcoR V
Not I
Xho I
BstE II*
Xba I*
Sac II*
V5 epitope
Age I
6xHis
stop
*Frame-dependent variations
Comments for pMIB/V5-His A
3596 nucleotides
OplE2 promoter: bases 1-549
OplE2 Forward priming site: bases 511-530
Honeybee melittin secretion signal: bases 565-627
Multiple cloning site: bases 629-721
V5 epitope: bases 734-775
Polyhistidine (6xHis) region: bases 785-802
OplE2 Reverse priming site: bases 812-837
OplE2 polyadenylation sequence: bases 820-949
pUC origin: bases 1018-1691 (complementary strand)
OplE1 promoter: bases 1765-2056
EM7 promoter: bases 2056-2122
Blasticidin resistance gene (bsd): bases 2123-2521
Ampicillin resistance gene (bla): bases 2641-3501
P OplE2
OplE2 pA
pUC ori
P OplE1
Blasticidin
Ampicillin
pMIB/V5-His
A, B, C
3.6 kb
B
pMIB/V5-His A MCS
Start of transcription
TATA Box
OplE2 Forward priming site
491 TCGCGCCTAT AAATACAGCC CGCAACGATC TGGTAAACAC AGTTGAACAG CATCTGTTCG AATTTAAAGC
Honeybee melittin secretion signal
561 TACC ATG AAA TTC TTA GTC AAC GTT GCC CTT GTT TTT ATG GTC GTA TAC ATT TCT
Met Lys Phe Leu Val Asn Val Ala Leu Val Phe Met Val Val Tyr Ile Ser
Sph I Hind III Asp718 I Kpn I Sac I BamH I Spe I EcoR I
616 TAC ATC TAT GCC GGC ATGCTAAGCT TGGTACCGAG CTCGGATCCA CTAGTCCAGT GTGGTGGAAT
Tyr Ile Tyr Ala
Melittin Cleavage Site
EcoR V Not I Xho I Xba I
681 TCTGCAGATA TCCAGCACAG TGGCGGCCGC TCGAGTCTAG AGGGCCCTTC GAA GGT AAG CCT ATC
Gly Lys Pro Ile
V5 epitope
Polyhistidine (6xHis) region
746 CCT AAC CCT CTC CTC GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC
Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg Thr Gly His His His His His
OplE2 Reverse priming site
800 CAT TGA GTTTA TCTGACTAAA TCTTAGTTTG TATTGTCATG TTTTAATACA ATATGTTATG
His ***
OplE2 polyadenylation signal
861 TTTAAATATG TTTTTAATAA ATTTTATAAA ATAATTTCAA CTTTTATTGT AACAACATTG TCCATTTACA
3' untranslated region of OplE2
931 CACTCCTTTC AAGCGCGTGG GATCGATGCT

Figure A4. **(A)** Map of the Insectselect™ pMIB/V5-His A-C protein expression vectors (Invitrogen). **(B)** Enlargement of the multiple cloning site of pMIB/V5-His A.

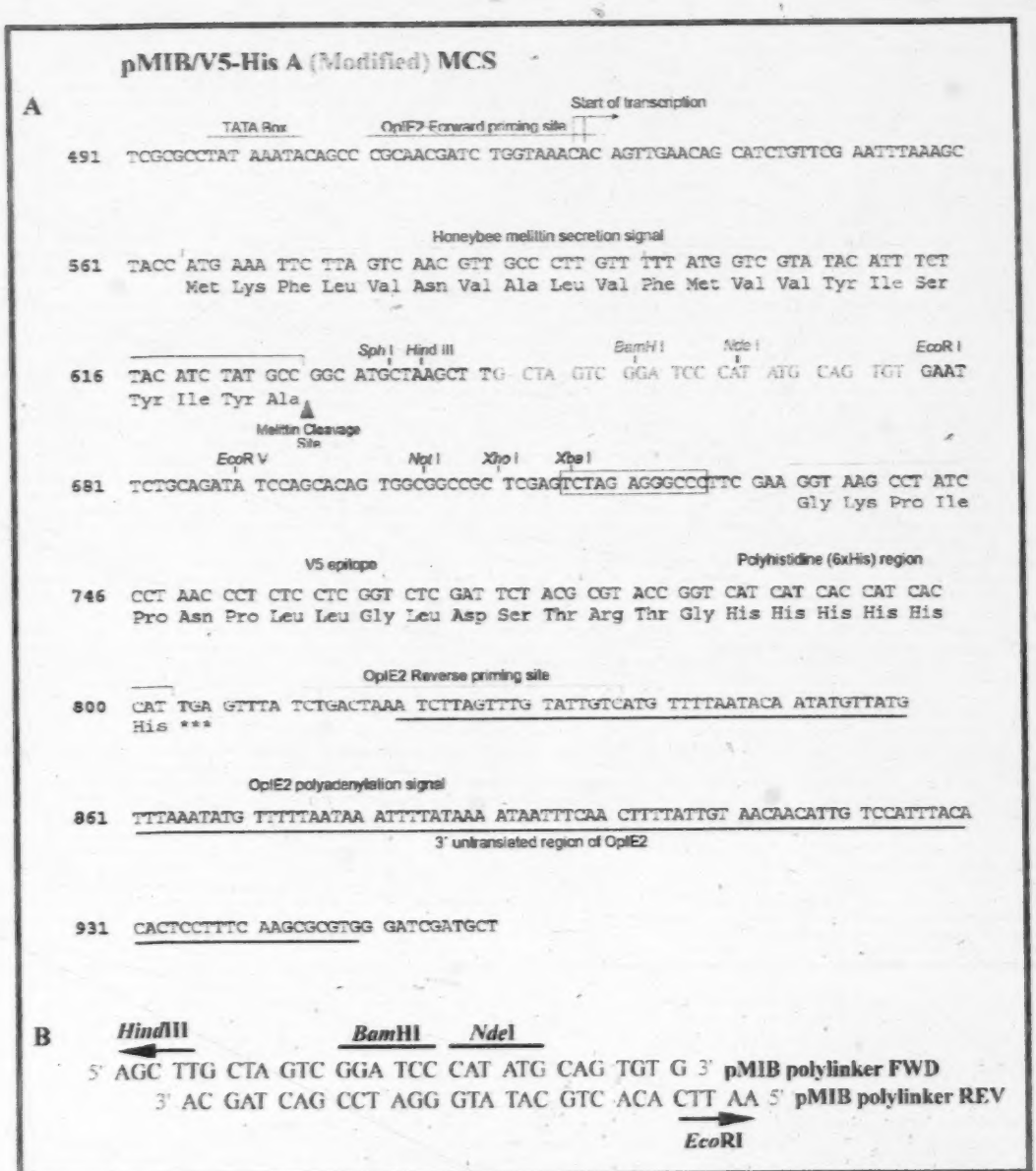

pMIB/V5-His A (Modified) MCS
A
Start of transcription
TATA Box
OpIE2 Forward priming site
491 TCGCGCCTAT AAATACAGCC CGCAACGATC TGGTAAACAC AGTTGAACAG CATCTGTTCG AATTTAAAGC
Honeybee melittin secretion signal
561 TACC ATG AAA TTC TTA GTC AAC GTT GCC CTT GTT TTT ATG GTC GTA TAC ATT TCT
Met Lys Phe Leu Val Asn Val Ala Leu Val Phe Met Val Val Tyr Ile Ser
Sph I Hind III
BamH I
Nde I
EcoR I
616 TAC ATC TAT GCC GGC ATGCTAAGCT TG CTA GTC GGA TCC CAT ATG CAG TGT GAAT
Tyr Ile Tyr Ala
Melittin Cleavage Site
EcoR V
Not I
Xho I
Xba I
681 TCTGCAGATA TCCAGCACAG TGGCGGCCGC TCGAGTCTAG AGGGCCCTTC GAA GGT AAG CCT ATC
Gly Lys Pro Ile
V5 epitope
Polyhistidine (6xHis) region
746 CCT AAC CCT CTC CTC GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC
Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg Thr Gly His His His His His
OpIE2 Reverse priming site
800 CAT TGA GTTTA TCTGACTAAA TCTTAGTTTG TATTGTCATG TTTTAATACA ATATGTTATG
His ***
OpIE2 polyadenylation signal
861 TTTAAATATG TTTTTAATAA ATTTTATAAA ATAATTTCAA CTTTTATTGT AACAACATTG TCCATTTACA
3' untranslated region of OpIE2
931 CACTCCTTTC AAGCGCGTGG GATCGATGCT
B
HindIII
BamHI
NdeI
5' AGC TTG CTA GTC GGA TCC CAT ATG CAG TGT G 3' pMIB polylinker FWD
3' AC GAT CAG CCT AGG GTA TAC GTC ACA CTT AA 5' pMIB polylinker REV
EcoRI

Figure A5. (A) The modified polylinker region of the Insectselect™ protein expression vector pMIB/V5-His A (Invitrogen). The portion in blue has been inserted to generate a *Bam*HI site in the +1 reading frame. **(B)** The two oligonucleotides used to insert into the *Hind*III/*Eco*RI cut pMIB/V5-His A vector to replace a portion of the polylinker.

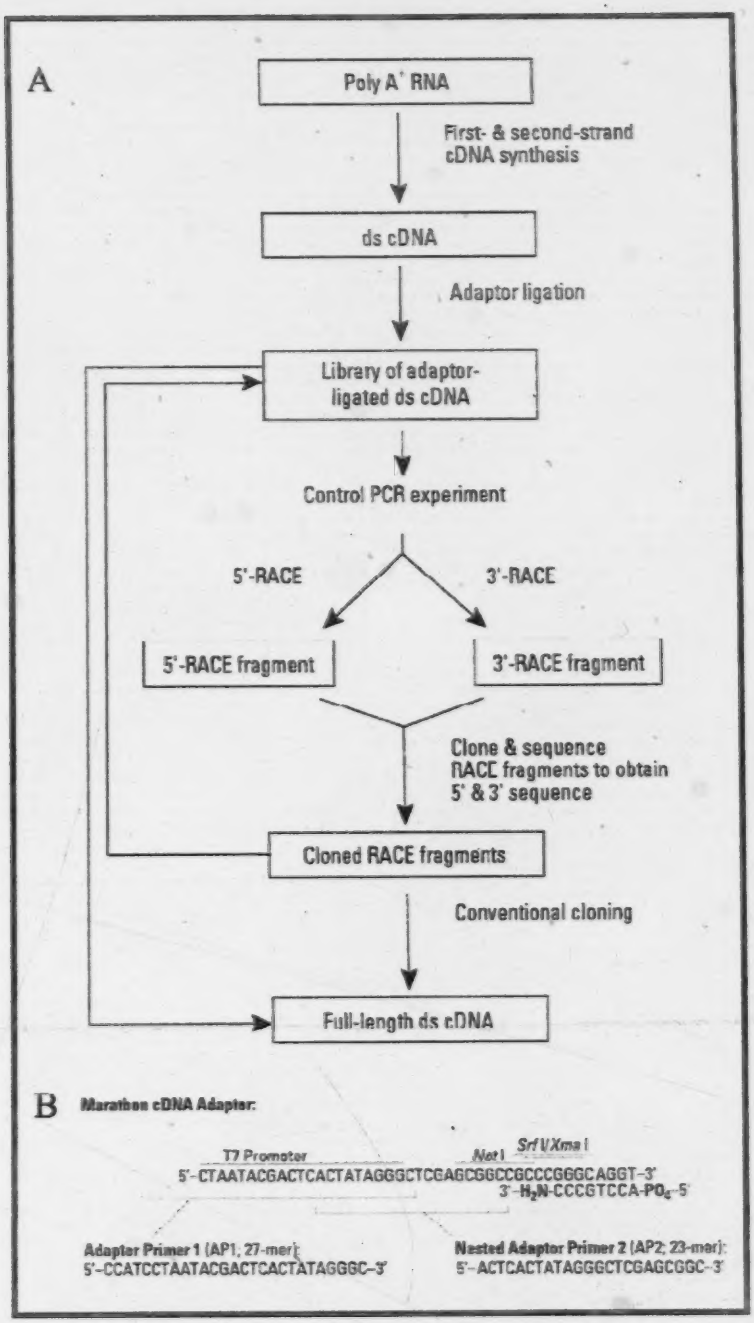


Figure A6. **(A)** Flow chart depicting the methodology employed for 5' RACE using Marathon™ adapter primers (Clontech). **(B)** Sequence of the Marathon™ adapter linker and the AP1 and AP2 linker specific primers.

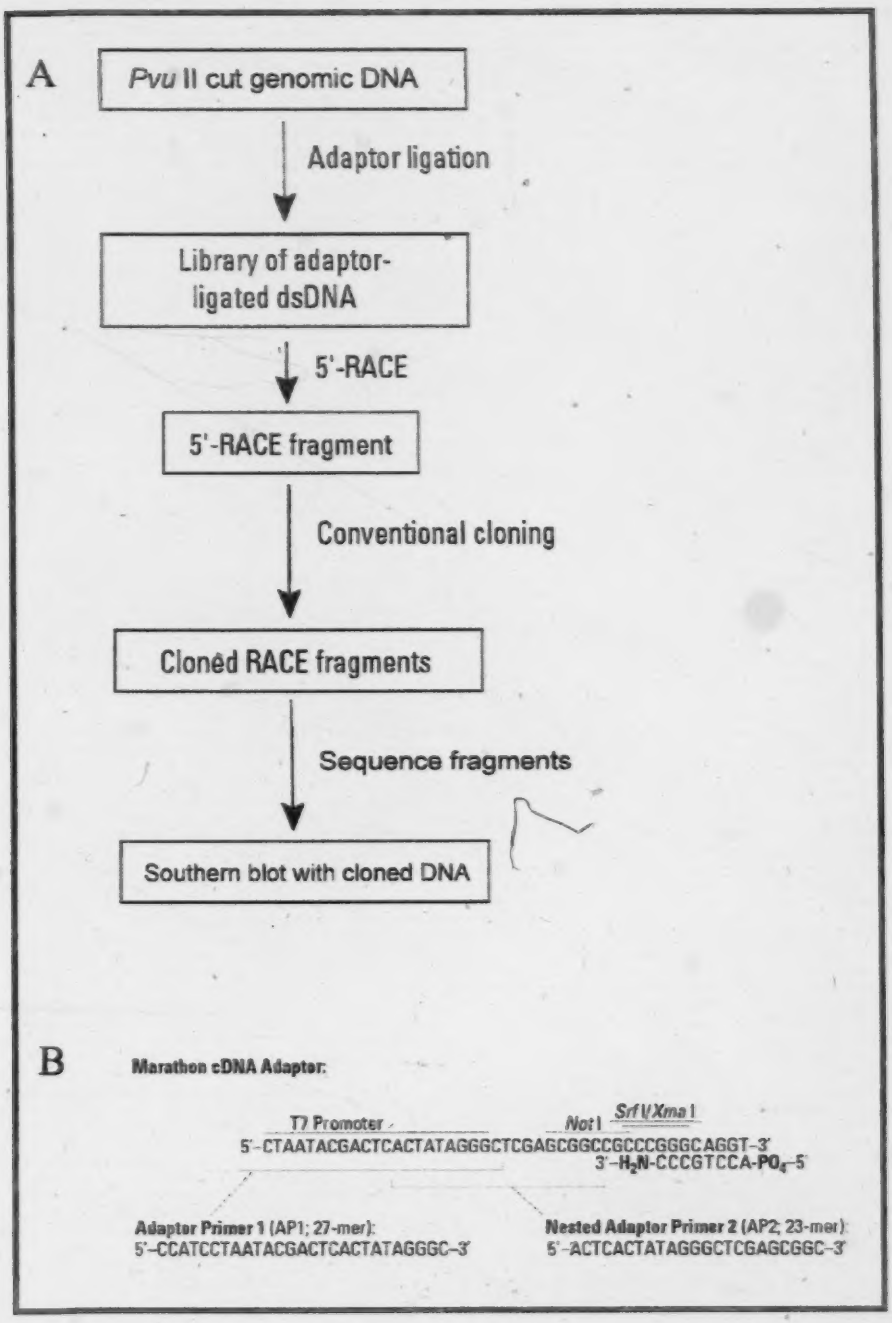


Figure A7. **(A)** Flow chart depicting the methodology employed for "RACE" of genomic DNA using Marathon™ adapter primers (Clontech). **(B)** Sequence of the Marathon™ adapter linker and the AP1 and AP2 linker specific primers.